

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2002, 03:04:37 ; Search time 11.3797 Seconds  
(without alignments)  
273.300 Million cell updates/sec

Title: US-09-847-539A-6\_COPY\_59\_86

Perfect score: 129

Sequence: 1 SDALALADQTDALQSEEAHVVKADNAA 28

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : A\_Geneseq\_032802.\*

```
1: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	129	100.0	28	21	AA1980
2	129	100.0	141	21	AA1981
3	129	100.0	155	21	AA1982
4	129	100.0	159	21	AA1983
5	129	100.0	167	21	AA1984
6	129	100.0	184	21	AA1985
7	129	100.0	217	21	AA1986
8	129	100.0	259	21	AA1987
9	129	100.0	271	21	AA1988
10	57	44.2	192	18	AA1989
11	57	44.2	222	18	AA1990

12	57	44.2	412	19	AA10994
13	50	38.8	1164	22	ABG30268
14	48.5	37.6	450	18	AAW11567
15	47	36.4	258	20	AA136771
16	46	35.7	705	22	AA190301
17	45.5	35.3	73	22	AB143356
18	45.5	35.3	73	22	AA164267
19	45.5	35.3	73	22	AA177089
20	45.5	35.3	73	22	AA172112
21	45.5	35.3	521	21	AA131518
22	45.5	35.3	572	21	AA131517
23	45.5	35.3	2076	22	AA134319
24	45.5	35.3	2186	22	AA137320
25	45	34.9	476	22	AA136317
26	45	34.9	502	21	AA154580
27	45	34.9	1362	20	AA199481
28	45	34.9	1447	20	AA181029
29	45	34.9	1637	22	AB160342
30	44.5	34.5	515	19	AA161907
31	44.5	34.5	516	19	AA170356
32	44.5	34.5	527	19	AA170357
33	44.5	34.5	552	19	AA161908
34	44	34.1	42	22	AB101077
35	44	34.1	274	21	AA158293
36	44	34.1	274	21	AA161028
37	44	34.1	284	22	AA178512
38	44	34.1	296	21	AA158292
39	44	34.1	296	21	AA161027
40	44	34.1	344	19	AA161234
41	44	34.1	358	21	AA158291
42	44	34.1	358	21	AA161026
43	44	34.1	371	21	AA181672
44	44	34.1	421	22	AB171817
45	44	34.1	429	22	AB165976

#### ALIGNMENTS

#### RESULT 1

AA171038  
ID AA171038 standard; peptide; 28 AA.

AC AA171038;

DT 29-AUG-2000 (first entry)

DE Streptococcus pyogenes strain SF370 GRAB protein fragment #3.

KW GRAB protein; protein G related alpha2M binding protein; vaccine;

KW alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;

KW immune response; Streptococcus pyogenes infection.

OS Streptococcus pyogenes.

PN WO200026240-A2.

PD 11-MAY-2000.

PF 02-NOV-1999; 99WO-GB03631.

PR 02-NOV-1999; 98GB-0023975.

PA (ACTI-) ACTINOVA LTD.

PI Björck LH, Rasmussen M;

DR WPI; 2000-365572/31.

PT New alpha2M binding protein for generating a protective immune response to group A streptococcus and purifying the binding protein

PS Claim 3; Page 56; 67pp; English.

H. pylori ORF 09ce  
Novel human diago  
E.coli alkaline ph  
Chlamydia trachoma  
C glutamicum prote  
Peptide #10862 enc  
Human bone marrow  
Human bone marrow  
Peptide #11249 enc  
Arabidopsis thalia  
Arabidopsis thalia  
Staphylococcus aur  
Staphylococcus aur  
Pseudomonas aerugi  
A rat calcium/calm  
Murine pCIP protei  
Murine pCIP protei  
Drosophila melanog  
Native RSP3 protei  
Chlamydomonas prot  
Tagged Chlamydomon  
Tandem tagged RSP3  
Novel human diago  
Arabidopsis thalia  
Arabidopsis thalia  
Human protein SPQ  
Arabidopsis thalia  
Arabidopsis thalia  
Streptococcus pneu  
Arabidopsis thalia  
Streptococcus pneu  
Arabidopsis thalia  
Drosophila melanog  
Drosophila melanog

XX The patent discloses a new family of proteins termed GRAB (protein G  
CC related alpha2M binding protein) from Streptococcus pyogenes which have  
CC the ability to bind alpha2-macroglobulin (alpha2M) and show homology to  
CC protein G of group G Streptococcus. GRAB protein and peptides derived  
CC from it are used in vaccine compositions for generating a protective  
CC immune response against group A Streptococcus. Antibodies against GRAB  
CC are useful for treating Streptococcus pyogenes infections. The protein  
CC is also useful for purifying alpha2M from a sample. The present sequence  
CC is one of the repeat regions of GRAB protein from  
CC S. pyogenes strain SF370 corresponding to residues 92-119. This fragment  
CC is useful in vaccine composition.  
XX  
SQ Sequence 28 AA;

Query Match 100.0%; Score 129; DB 21; Length 28;  
Best Local Similarity 100.0%; Pred. No. 1.4e-11;  
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SDALALADQTDALQSEEAAYVVKADNAA 28  
|||||  
DB 1 SDALALADQTDALQSEEAAYVVKADNAA 28

RESULT 2  
AAV71040  
ID AAV71040 standard; peptide; 141 AA.  
XX  
AC AAV71040;  
XX  
DT 29-AUG-2000 (first entry)  
XX  
DE Streptococcus pyogenes strain SF370 GRAB protein fragment #4.  
XX  
KW GRAB protein; protein G related alpha2M binding protein; vaccine;  
KW alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;  
KW immune response; Streptococcus pyogenes infection.  
XX  
OS Streptococcus pyogenes.  
XX  
PN WO200026240-A2.  
XX  
PD 11-MAY-2000.  
XX  
PF 02-NOV-1999; 99WO-GB03631.  
XX  
PR 02-NOV-1998; 98GB-0023975.  
XX  
PA (ACTI-) ACTINOVA LTD.  
XX  
PI Bjorck LH, Rasmussen M;  
XX  
DR WPI; 2000-365572/31.  
XX  
PR 02-NOV-1998; 98GB-0023975.  
XX  
PA (ACTI-) ACTINOVA LTD.  
XX  
PI Bjorck LH, Rasmussen M;  
XX  
DR WPI; 2000-365572/31.  
XX

XX New alpha2M binding protein for generating a protective immune response  
PT to group A streptococcus and purifying the binding protein -  
XX  
PS Claim 5; Page 57; 67pp; English.  
XX

XX The patent discloses a new family of proteins termed GRAB (protein G  
CC related alpha2M binding protein) from Streptococcus pyogenes which have  
CC the ability to bind alpha2-macroglobulin (alpha2M) and show homology to  
CC protein G of group G Streptococcus. GRAB protein and peptides derived  
CC from it are used in vaccine compositions for generating a protective  
CC immune response against group A Streptococcus. Antibodies against GRAB  
CC are useful for treating Streptococcus pyogenes infections. The protein  
CC is also useful for purifying alpha2M from a sample. The present sequence  
CC is a fragment of GRAB protein from S. pyogenes strain SF370 corresponding  
CC to residues 34-174. This fragment is devoid of the trans-membrane  
CC and cell wall anchor regions. It is useful in vaccine composition.  
XX  
SQ Sequence 141 AA;

Query Match 100.0%; Score 129; DB 21; Length 141;  
Best Local Similarity 100.0%; Pred. No. 7.6e-11;  
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SDALALADQTDALQSEEAAYVVKADNAA 28  
|||||  
DB 59 SDALALADQTDALQSEEAAYVVKADNAA 86

RESULT 3  
AAV71044  
ID AAV71044 standard; Protein; 155 AA.  
XX  
AC AAV71044;  
XX

XX 29-AUG-2000 (first entry)  
XX  
DE Streptococcus pyogenes strain AP1 partial GRAB protein.  
XX

KW GRAB protein; protein G related alpha2M binding protein; vaccine;  
KW alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;  
KW immune response; Streptococcus pyogenes infection.  
XX  
OS Streptococcus pyogenes.  
XX

PH Key Location/Qualifiers  
FT Misc-difference 17  
FT /note= "Thr at position 18 of GRAB protein fragment  
FT (AAV71036) from S. pyogenes strain SF370 is replaced  
FT with Ile"  
XX  
PN WO200026240-A2.  
XX  
PD 11-MAY-2000.  
XX  
PF 02-NOV-1999; 99WO-GB03631.  
XX  
PR 02-NOV-1998; 98GB-0023975.  
XX  
PA (ACTI-) ACTINOVA LTD.  
XX  
PI Bjorck LH, Rasmussen M;  
XX  
DR WPI; 2000-365572/31.  
XX  
PR N-PSDB; AAD00562.  
XX

PT New alpha2M binding protein for generating a protective immune response  
PT to group A streptococcus and purifying the binding protein -  
XX  
PS Claim 5; Page 60-61; 67pp; English.  
XX

XX The patent discloses a new family of proteins termed GRAB (protein G  
CC related alpha2M binding protein) from Streptococcus pyogenes which have  
CC the ability to bind alpha2-macroglobulin (alpha2M) and show homology to  
CC protein G of group G Streptococcus. GRAB protein and peptides derived  
CC from it are used in vaccine compositions for generating a protective  
CC immune response against group A Streptococcus. Antibodies against GRAB  
CC are useful for treating Streptococcus pyogenes infections. The protein  
CC is also useful for purifying alpha2M from a sample. The present sequence  
CC is a partial GRAB protein from S. pyogenes strain AP1.  
CC The protein has alpha2M binding region and is useful in vaccine  
CC composition.  
XX  
SQ Sequence 155 AA;

Query Match 100.0%; Score 129; DB 21; Length 155;  
Best Local Similarity 100.0%; Pred. No. 8.4e-11;  
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SDALALADQTDALQSEEAAYVVKADNAA 28  
|||||  
DB 58 SDALALADQTDALQSEEAAYVVKADNAA 85



```
RESULT 4
AAY71041
ID AAY71041 standard; peptide; 159 AA.
XX
XX
AC AAY71041;
XX
XX 29-AUG-2000 (first entry)
DT
XX
XX Streptococcus pyogenes strain SF370 GRAB protein fragment #5.
DE
XX
XX GRAB protein; protein G related alpha2M binding protein; vaccine;
KW alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;
KW immune response; Streptococcus pyogenes infection.
XX
OS Streptococcus pyogenes.
XX
PN WO200026240-A2.
XX
PD 11-MAY-2000.
XX
XX 02-NOV-1999; 99WO-GB03631.
PF
XX 02-NOV-1998; 98GB-0023975.
PR
XX (ACTI-) ACTINOVA LTD.
PA
PI Bjorck LH, Rasmussen M;
XX
XX WPI; 2000-365572/31.
DR
XX N-PSDB; AAD00564.
XX
XX New alpha2M binding protein for generating a protective immune response
PT to group A streptococcus and purifying the binding protein -
XX
XX Claim 5; Page 57-58; 67pp; English.
XX
XX The patent discloses a new family of proteins termed GRAB (protein G
CC related alpha2M binding protein) from Streptococcus pyogenes which have
CC the ability to bind alpha2-macroglobulin (alpha2M) and show homology to
CC protein G of group G Streptococcus. GRAB protein and peptides derived
CC from it are used in vaccine compositions for generating a protective
CC immune response against group A Streptococcus. Antibodies against GRAB
CC are useful for treating Streptococcus pyogenes infections. The protein
CC is also useful for purifying alpha2M from a sample. The present sequence
CC is a fragment of GRAB protein from S. pyogenes strain SF370 corresponding
CC to residues 34-192 and devoid of the membrane spanning region. This
CC fragment is useful in vaccine composition.
XX
XX Query Match 100.0%; Score 129; DB 21; Length 159;
XX Best Local Similarity 100.0%; Pred. No. 8.7e-11;
XX Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
OY 1 SDALEALADQTDALQSEEAAYVKADNAA 28
DB 59 SDALEALADQTDALQSEEAAYVKADNAA 86
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
RESULT 5
AAY71046
ID AAY71046 standard; Protein; 167 AA.
XX
XX
AC AAY71046;
XX
XX 29-AUG-2000 (first entry)
DT
XX
XX Streptococcus pyogenes strain KTL3 partial GRAB protein.
DE
XX
XX GRAB protein; protein G related alpha2M binding protein; vaccine;
KW alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;
KW immune response; Streptococcus pyogenes infection.
XX
```

```
OS Streptococcus pyogenes.
XX
XX WO200026240-A2.
XX
XX 11-MAY-2000.
XX
XX 02-NOV-1999; 99WO-GB03631.
PF
XX 02-NOV-1998; 98GB-0023975.
PR
XX (ACTI-) ACTINOVA LTD.
PA
PI Bjorck LH, Rasmussen M;
XX
XX WPI; 2000-365572/31.
DR
XX N-PSDB; AAD00564.
XX
XX New alpha2M binding protein for generating a protective immune response
PT to group A streptococcus and purifying the binding protein -
XX
XX Claim 5; Page 62-63; 67pp; English.
XX
XX The patent discloses a new family of proteins termed GRAB (protein G
CC related alpha2M binding protein) from Streptococcus pyogenes which have
CC the ability to bind alpha2-macroglobulin (alpha2M) and show homology to
CC protein G of group G Streptococcus. GRAB protein and peptides derived
CC from it are used in vaccine compositions for generating a protective
CC immune response against group A Streptococcus. Antibodies against GRAB
CC are useful for treating Streptococcus pyogenes infections. The protein
CC is also useful for purifying alpha2M from a sample. The present sequence
CC is a partial GRAB protein from S. pyogenes strain KTL3.
CC The protein has alpha2M binding region and is useful in vaccine
CC composition.
XX
XX Query Match 100.0%; Score 129; DB 21; Length 167;
XX Best Local Similarity 100.0%; Pred. No. 9.2e-11;
XX Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
OY 1 SDALEALADQTDALQSEEAAYVKADNAA 28
DB 70 SDALEALADQTDALQSEEAAYVKADNAA 97
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
RESULT 6
AAY71039
ID AAY71039 standard; Protein; 184 AA.
XX
XX
AC AAY71039;
XX
XX 29-AUG-2000 (first entry)
DT
XX
XX Streptococcus pyogenes strain SF370 mature GRAB protein.
DE
XX
XX GRAB protein; protein G related alpha2M binding protein; vaccine;
KW alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;
KW immune response; Streptococcus pyogenes infection.
XX
XX Streptococcus pyogenes.
OS
XX
XX Key Location/Qualifiers
FH Binding-site 1..58
FT /label= alpha2-macroglobulin_binding_site
FT Region 59..86
FT /label= R1
FT /note= "repeat region"
FT Region 87..114
FT /label= R2
FT /note= "repeat region"
FT Region 115..159
FT /label= Cell_wall_spanning_region
FT Region 150..155
```

```
FT FT /note= "consensus sequence for gram-positive
FT FT surface cell wall anchored proteins"
FT FT 160..184
FT FT /label= Membrane_spanning_region
XX
PN WO200026240-A2.
XX
XX 11-MAY-2000.
XX
XX 02-NOV-1999; 99WO-GB03631.
XX
XX 02-NOV-1998; 98GB-0023975.
XX
XX (ACTI-) ACTINOVA LTD.
XX
XX Bjorck LH, Rasmussen M;
XX
XX WPI; 2000-365572/31.
XX
XX New alpha2M binding protein for generating a protective immune response
XX to group A streptococcus and purifying the binding protein -
XX
XX Claim 5; Page 56; 67pp; English.
XX
XX The patent discloses a new family of proteins termed GRAB (protein G
XX related alpha2M binding protein) from Streptococcus pyogenes which have
XX the ability to bind alpha2-macroglobulin (alpha2M) and show homology to
XX protein G of group G Streptococcus. GRAB protein and peptides derived
XX from it are used in vaccine compositions for generating a protective
XX immune response against group A Streptococcus. Antibodies against GRAB
XX are useful for treating Streptococcus pyogenes infections. The protein
XX is a mature GRAB protein from S. pyogenes strain SF370 without the
XX signal sequence. This sequence is capable of binding alpha2M
XX and useful in vaccine composition.
XX
XX Sequence 184 AA;
XX
XX Query Match 100.0%; Score 129; DB 21; Length 184;
XX Best Local Similarity 100.0%; Pred. No. 1e-10;
XX Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX Qy 1 SDALEALADQTDALQSEEAAYVVKADNAA 28
XX |||||||
XX Db 59 SDALEALADQTDALQSEEAAYVVKADNAA 86
XX
XX RESULT 7
XX AAY71042
XX ID AAY71042 standard; Protein: 217 AA.
XX
XX AC AAY71042;
XX
XX DT 29-AUG-2000 (first entry)
XX
XX DE Streptococcus pyogenes strain SF370 full-length GRAB protein.
XX
XX KW GRAB protein; protein G related alpha2M binding protein; vaccine;
XX alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;
XX immune response; Streptococcus pyogenes infection.
XX
XX OS Streptococcus pyogenes.
XX
XX FH Key Location/Qualifiers
XX FT Peptide 1..33
XX FT /label= Signal_sequence
XX FT Protein 34..217
XX FT /label= Mature_GRAB_protein
XX FT Binding-site 34..91
XX FT /label= alpha2M_binding_site
XX FT Domain 34..68
XX FT /note= "shows homology to E domain of protein G"
XX FT Region 92..119
```

```
FT Region /label= Repeat_region_1
FT 120..147
FT /label= Repeat_region_2
FT 148..192
FT /label= Cell_wall_spanning_region
FT 183..188
FT /note= "consensus sequence for gram-positive
FT surface cell wall anchored proteins"
FT 193..217
FT /label= Membrane_spanning_region
XX
XX WO200026240-A2.
XX
XX 11-MAY-2000.
XX
XX 02-NOV-1999; 99WO-GB03631.
XX
XX 02-NOV-1998; 98GB-0023975.
XX
XX (ACTI-) ACTINOVA LTD.
XX
XX Bjorck LH, Rasmussen M;
XX
XX WPI; 2000-365572/31.
XX
XX N-PSDB; AAD00559; AAD00560.
XX
XX New alpha2M binding protein for generating a protective immune response
XX to group A streptococcus and purifying the binding protein -
XX
XX Claim 5; Fig 2B; 67pp; English.
XX
XX The patent discloses a new family of proteins termed GRAB (protein G
XX related alpha2M binding protein) from Streptococcus pyogenes which have
XX the ability to bind alpha2-macroglobulin (alpha2M) and show homology to
XX protein G of group G Streptococcus. GRAB protein and peptides derived
XX from it are used in vaccine compositions for generating a protective
XX immune response against group A Streptococcus. Antibodies against GRAB
XX are useful for treating Streptococcus pyogenes infections. The protein
XX is also useful for purifying alpha2M from a sample. The present sequence
XX is a full-length GRAB protein from S. pyogenes strain SF370.
XX
XX Sequence 217 AA;
XX
XX Query Match 100.0%; Score 129; DB 21; Length 217;
XX Best Local Similarity 100.0%; Pred. No. 1.2e-10;
XX Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX Qy 1 SDALEALADQTDALQSEEAAYVVKADNAA 28
XX |||||||
XX Db 92 SDALEALADQTDALQSEEAAYVVKADNAA 119
XX
XX RESULT 8
XX AAY71043
XX ID AAY71043 standard; Protein: 259 AA.
XX
XX AC AAY71043;
XX
XX DT 29-AUG-2000 (first entry)
XX
XX DE Streptococcus pyogenes strain KTL9 partial GRAB protein.
XX
XX KW GRAB protein; protein G related alpha2M binding protein; vaccine;
XX alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;
XX immune response; Streptococcus pyogenes infection.
XX
XX OS Streptococcus pyogenes.
XX
XX PN WO200026240-A2.
XX
XX PD 11-MAY-2000.
XX
XX 02-NOV-1999; 99WO-GB03631.
```





OY 4 LEALADQTDALQSEAAVVKADN 26  
: : : : : : : : : : : : : : : :  
Db 49 IQALQEOIDALDSQEKVSKWDN 71

RESULT 13  
ABG30268  
ID ABG30268 standard; Protein; 1164 AA.

XX ABG30268;  
XX AC  
XX DT 18-FEB-2002 (first entry)  
XX DE Novel human diagnostic protein #30259.  
XX DE Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
KW OS Homo sapiens.  
XX WO200175067-A2.  
XX PD 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.  
XX PF 31-MAR-2000; 2000US-0540217.  
XX PR 23-AUG-2000; 2000US-0649167.  
XX (HYSE-) HYSEQ INC.

XX Dmanac RT, Liu C, Tang YT;  
XX WPI: 2001-639362/73.  
XX N-PSDB; AAS94455.

XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity

XX Claim 20; SEQ ID No 60627; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and  
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 1164 AA;

Query Match 38.8%; Score 50; DB 22; Length 1164;  
Best Local Similarity 45.8%; Pred. No. 1.5e+02;  
Matches 11; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

OY 3 ALEALADQTDALQSEAAVVKADN 26  
: : : : : : : : : : : : : : : :  
XX

Db 581 AVIQLTPQTEAVRSEDAPVAPRDN 604

RESULT 14  
AAW11567  
ID AAW11567 standard; Protein; 450 AA.

XX AAW11567;  
XX DT 20-OCT-1997 (first entry)  
XX DE E.coli alkaline phosphatase mutant Q329A.  
XX KW Bacterial alkaline phosphatase; BAP; mutein; variant; inactive;  
KW immunoassay.  
XX OS Escherichia coli.  
XX OS Synthetic.

XX Key Location/Qualifiers  
FH Misc-difference 330 /note= "Wild-type Gln329 is replaced by Ala"  
FT

XX EP752475-A1.

XX 08-JAN-1997.

XX 27-JUN-1996; 96EP-0401419.

XX 29-JUN-1995; 95PR-0007833.

XX (BOUL/) BOULAIN J.  
XX (COMS ) COMMISSARIAT ENERGIE ATOMIQUE.

XX Boulain J, Cattolico L, Ducancel F, Menez A;

XX WPI: 1997-067452/07.

XX Modified bacterial alkaline phosphatase enzymes - useful as  
PT immunoassay labels

XX Claim 6; Page -: 35pp; French.

XX This sequence represents a specifically claimed modified bacterial  
CC alkaline phosphatase (BAP) that has higher activity and similar  
CC thermal stability compared with the corresponding wild-type E.coli  
CC BAP. In this mutein, one wild-type amino acid residue has been  
CC substituted (see features table). Optionally, at least one amino acid  
CC is inserted between residues 6 (Pro) and 7 (Val) of the wild-type  
CC enzyme. The modified enzyme is useful as an immunoassay label.

CC Note: This sequence does not appear in the specification but has  
CC been derived by modifying an E.coli BAP sequence according to the  
CC numbering given in the sequence listing.

XX Sequence 450 AA;

Query Match 37.6%; Score 48.5; DB 18; Length 450;  
Best Local Similarity 34.1%; Pred. No. 80;  
Matches 14; Conservative 6; Mismatches 8; Indels 13; Gaps 1;

OY 1 SDALCALADQTD-----LQSEAAVVKADNAA 28  
: : : : : : : : : : : : : : : :  
Db 294 NDSVPTLAQMTDKAIELLSKNEKGFQFLOVEGASIDKADHAA 334

RESULT 15  
AAY36771  
ID AAY36771 standard; Protein; 258 AA.

XX AAY36771;

XX DT 07-OCT-1999 (first entry)



```
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483447/52.
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human fetal liver -
XX Claim 27; SEQ ID NO 35991; 639pp + sequence listing; English.
XX The invention relates to a single exon nucleic acid probe for
XX measuring human gene expression in a sample derived from human foetal
XX liver. The single exon nucleic acid probes may be used for predicting,
XX measuring and displaying gene expression in samples derived from human
XX fetal liver. The present sequence is a peptide encoded by a single exon
XX nucleic acid probe of the invention.
XX Note: The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 73 AA;
XX
XX Query Match 35.3%; Score 45.5; DB 22; Length 73;
XX Best Local Similarity 48.1%; Pred. No. 26;
XX Matches 13; Conservative 2; Mismatches 7; Indels 5; Gaps 1;
XX
XX QY 2 DALEALADQTDALQS-----EEAAVVK 23
XX | | | | | | | | | | | | | | | | | |
XX Db 47 DILEQTAAGVDALQGHLMQMEQALVK 73
XX
XX RESULT 18
XX AAM64267
XX ID AAM64267 standard; Protein; 73 AA.
XX AC AAM64267;
XX XX
XX DT 05-NOV-2001 (first entry)
XX DE Human brain expressed single exon probe encoded protein SEQ ID NO: 36372.
XX KW Human; brain expressed exon; gene expression analysis; probe;
XX KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
XX KW epilepsy; cancer.
XX OS Homo sapiens.
XX PN WO200157275-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00667.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX DR WPI; 2001-483446/52.
XX XX
XX PT Single exon nucleic acid probes for analyzing gene expression in human
XX brains -
XX PS Example 4; SEQ ID NO: 36372; 650pp + Sequence Listing; English.
XX XX
XX CC The present invention provides a number of single exon nucleic acid
```

```
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia, f
CC epilepsy and cancers. The present sequence is a protein encoded by one of
CC the probes of the invention.
XX Sequence 73 AA;
XX
XX Query Match 35.3%; Score 45.5; DB 22; Length 73;
XX Best Local Similarity 48.1%; Pred. No. 26;
XX Matches 13; Conservative 2; Mismatches 7; Indels 5; Gaps 1;
XX
XX QY 2 DALEALADQTDALQS-----EEAAVVK 23
XX | | | | | | | | | | | | | | | | | |
XX Db 47 DILEQTAAGVDALQGHLMQMEQALVK 73
XX
XX RESULT 19
XX AAM77089
XX ID AAM77089 standard; Protein; 73 AA.
XX AC AAM77089;
XX XX
XX DT 06-NOV-2001 (first entry)
XX DE Human bone marrow expressed probe encoded protein SEQ ID NO: 37395.
XX KW Human; bone marrow expressed exon; gene expression analysis; probe;
XX KW microarray; cancer; leukaemia; lymphoma; myeloma.
XX OS Homo sapiens.
XX PN WO200157276-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00668.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX DR WPI; 2001-488900/53.
XX XX
XX PT Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human bone marrow -
XX PS Example 4; SEQ ID NO: 37395; 658pp + Sequence Listing; English.
XX XX
XX CC The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX bone marrow. They can be used to measure gene expression in bone marrow
XX samples, which may enable the improved diagnosis and treatment of cancers
XX such as lymphoma, leukaemia and myeloma. The present sequence is a
XX protein encoded by one of the probes of the invention.
XX Sequence 73 AA;
XX
XX Query Match 35.3%; Score 45.5; DB 22; Length 73;
XX Best Local Similarity 48.1%; Pred. No. 26;
XX Matches 13; Conservative 2; Mismatches 7; Indels 5; Gaps 1;
XX
XX QY 2 DALEALADQTDALQS-----EEAAVVK 23
XX | | | | | | | | | | | | | | | | | |
```





PR	28-JUN-1999;	99US-0140823;
PR	29-JUN-1999;	99US-0140991;
PR	30-JUN-1999;	99US-0141287;
PR	01-JUL-1999;	99US-0141842;
PR	01-JUL-1999;	99US-0142154;
PR	02-JUL-1999;	99US-0142055;
PR	08-JUL-1999;	99US-0142390;
PR	08-JUL-1999;	99US-0142803;
PR	09-JUL-1999;	99US-0142970;
PR	12-JUL-1999;	99US-0142927;
PR	13-JUL-1999;	99US-0143542;
PR	14-JUL-1999;	99US-0143624;
PR	15-JUL-1999;	99US-0144005;
PR	16-JUL-1999;	99US-0144085;
PR	16-JUL-1999;	99US-0144086;
PR	19-JUL-1999;	99US-0144325;
PR	19-JUL-1999;	99US-0144331;
PR	19-JUL-1999;	99US-0144332;
PR	19-JUL-1999;	99US-0144333;
PR	19-JUL-1999;	99US-0144334;
PR	19-JUL-1999;	99US-0144335;
PR	20-JUL-1999;	99US-0144352;
PR	20-JUL-1999;	99US-0144632;
PR	20-JUL-1999;	99US-0144884;
PR	21-JUL-1999;	99US-0144814;
PR	21-JUL-1999;	99US-0145086;
PR	21-JUL-1999;	99US-0145088;
PR	22-JUL-1999;	99US-0145085;
PR	22-JUL-1999;	99US-0145087;
PR	22-JUL-1999;	99US-0145089;
PR	22-JUL-1999;	99US-0145192;
PR	23-JUL-1999;	99US-0145145;
PR	23-JUL-1999;	99US-0145218;
PR	23-JUL-1999;	99US-0145224;
PR	26-JUL-1999;	99US-0145276;
PR	27-JUL-1999;	99US-0145913;
PR	27-JUL-1999;	99US-0145918;
PR	27-JUL-1999;	99US-0145919;
PR	28-JUL-1999;	99US-0145951;
PR	02-AUG-1999;	99US-0146386;
PR	02-AUG-1999;	99US-0146388;
PR	02-AUG-1999;	99US-0146389;
PR	03-AUG-1999;	99US-0147038;
PR	04-AUG-1999;	99US-0147204;
PR	04-AUG-1999;	99US-0147302;
PR	05-AUG-1999;	99US-0147192;
PR	05-AUG-1999;	99US-0147260;
PR	06-AUG-1999;	99US-0147303;
PR	06-AUG-1999;	99US-0147416;
PR	09-AUG-1999;	99US-0147493;
PR	09-AUG-1999;	99US-0147935;
PR	10-AUG-1999;	99US-0148171;
PR	11-AUG-1999;	99US-0148319;
PR	12-AUG-1999;	99US-0148341;
PR	13-AUG-1999;	99US-0148565;
PR	13-AUG-1999;	99US-0148684;
PR	16-AUG-1999;	99US-0149368;
PR	17-AUG-1999;	99US-0149175;
PR	18-AUG-1999;	99US-0149426;
PR	20-AUG-1999;	99US-0149723;
PR	20-AUG-1999;	99US-0149929;
PR	23-AUG-1999;	99US-0149902;
PR	23-AUG-1999;	99US-0149930;
PR	25-AUG-1999;	99US-0150566;
PR	26-AUG-1999;	99US-0150884;
PR	27-AUG-1999;	99US-0151065;
PR	27-AUG-1999;	99US-0151066;
PR	27-AUG-1999;	99US-0151080;
PR	30-AUG-1999;	99US-0151303;
PR	31-AUG-1999;	99US-0151436;
PR	01-SEP-1999;	99US-0151930;
PR	07-SEP-1999;	99US-0152363;

PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.
PR	13-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.

Query Match 35.3%; Score 45.5; DB 21; Length 521;  
Best Local Similarity 61.9%; Pred. No. 2.5e+02;  
Matches 13; Conservative 2; Mismatches 5; Indels 1

Qy 4 LEALADQTDALQSEEAUVKA 24  
| | | | : | | | | |  
pb 40 LEALASET-LLTSDEAAVVAA 59

RESULT 22  
AAG31517

AAG31517  
ID AAG31517 standard: Protein: 572 AA:

XX AC AAG31517:

XX  
DT 17-OCT-2000 (first entry)

XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 37862.

XX Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.

XX Arabidopsis thaliana.

XX  
PN EP1033405-A2.

XX  
PD 06-SEP-2000



PR 24-SEP-1999; 99US-0155659.  
 PR 28-SEP-1999; 99US-0156458.  
 PR 29-SEP-1999; 99US-0156596.  
 PR 04-OCT-1999; 99US-0157117.  
 PR 05-OCT-1999; 99US-0157753.  
 PR 06-OCT-1999; 99US-0157865.  
 PR 07-OCT-1999; 99US-0158029.  
 PR 08-OCT-1999; 99US-0158232.  
 PR 12-OCT-1999; 99US-0158369.  
 PR 13-OCT-1999; 99US-0159293.  
 PR 13-OCT-1999; 99US-0159294.  
 PR 14-OCT-1999; 99US-0159295.  
 PR 14-OCT-1999; 99US-0159329.  
 PR 14-OCT-1999; 99US-0159330.  
 PR 14-OCT-1999; 99US-0159331.  
 PR 14-OCT-1999; 99US-0159637.  
 PR 18-OCT-1999; 99US-0159638.  
 PR 21-OCT-1999; 99US-0159584.  
 PR 21-OCT-1999; 99US-0160741.  
 PR 21-OCT-1999; 99US-0160767.  
 PR 21-OCT-1999; 99US-0160768.  
 PR 21-OCT-1999; 99US-0160770.  
 PR 21-OCT-1999; 99US-0160814.  
 PR 21-OCT-1999; 99US-0160815.  
 PR 22-OCT-1999; 99US-0160980.  
 PR 22-OCT-1999; 99US-0160981.  
 PR 22-OCT-1999; 99US-0160989.  
 PR 25-OCT-1999; 99US-0161404.  
 PR 25-OCT-1999; 99US-0161405.  
 PR 25-OCT-1999; 99US-0161406.  
 PR 26-OCT-1999; 99US-0161359.  
 PR 26-OCT-1999; 99US-0161360.  
 PR 26-OCT-1999; 99US-0161361.  
 PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161952.  
 PR 28-OCT-1999; 99US-0161993.  
 PR 29-OCT-1999; 99US-0162142.

Query Match 35.3%; Score 45.5; DB 21; Length 572;  
 Best Local Similarity 61.9%; Pred. No. 2.8e+02;  
 Matches 13; Conservative 2; Mismatches 5; Indels 1; Gaps 1;

QY 4 LEALADQTDALQSEEAHVKA 24  
 ||||:| | | | | | | |  
 Db 91 LEALASET-LLTSDEAVVAA 110

RESULT 23  
 AAU34319  
 ID AAU34319 standard; Protein; 2076 AA.

XX AC AAU34319;

XX DT 14-FEB-2002 (first entry)

XX DE Staphylococcus aureus cellular proliferation protein #595.

XX KW Antisense; prokaryotic cellular proliferation protein;

XX KW antibiotic; antibacterial; drug design.

XX .OS Staphylococcus aureus.

XX PN WO200170955-A2.

XX PD 27-SEP-2001.

XX XX 21-MAR-2001; 2001WO-US09180.

XX XX 21-MAR-2000; 2000US-191078P.

PR 23-MAY-2000; 2000US-206848P.

PR 26-MAY-2000; 2000US-207727P.

PR 23-OCT-2000; 2000US-242578P.

PR 27-NOV-2000; 2000US-253625P.

PR 22-DEC-2000; 2000US-257931P.  
 PR 16-FEB-2001; 2001US-269308P.  
 XX PA (ELIT-) ELITRA PHARM INC.

XX PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;  
 PI Yamamoto RT, Xu HH;  
 XX N-PSDB; AAS521178.

DR WPI; 2001-611495/70.

XX New polynucleotides for the identification and development of

PT antibiotics, comprise sequences of antisense nucleic acids -

XX Example 3; Seq ID No 5815; 511pp; English.

CC The invention relates to antisense inhibitors of genes essential to  
 CC prokaryotic cellular proliferation, their use in identifying the  
 CC genes, their use in the discovery of novel antibiotics, the essential  
 CC genes themselves and the encoded proteins. The prokaryotes used are  
 CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella  
 CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The  
 CC invention is also useful for the identification of potential new targets  
 CC for antibiotic development. The antisense nucleic acids can also be used  
 CC to identify proteins used in proliferation, to express these proteins,  
 CC and to obtain antibodies capable of binding to the expressed proteins.  
 CC The proteins can be used to screen compounds in rational drug discovery  
 CC programmes. The antisense nucleic acid sequence is also useful to screen  
 CC for homologous nucleic acids which are required for cell proliferation in  
 CC a wide variety of organisms. The present sequence represents an  
 CC essential prokaryotic cellular proliferation protein.

CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 2076 AA;

Query Match 35.3%; Score 45.5; DB 22; Length 2076;  
 Best Local Similarity 48.0%; Pred. No. 1.3e+03;  
 Matches 12; Conservative 3; Mismatches 9; Indels 1; Gaps 1;

QY 2 DALEALA-DQTDALQSEEAHVKA 25  
 ||| | | | | | | | | |  
 Db 766 DALNQLATDETDIDVNTNATNAD 790

RESULT 24  
 AAU37320  
 ID AAU37320 standard; Protein; 2186 AA.

XX AC AAU37320;

XX DT 14-FEB-2002 (first entry)

XX DE Staphylococcus aureus cellular proliferation protein #1490.

XX KW Antisense; prokaryotic cellular proliferation protein;

XX KW antibiotic; antibacterial; drug design.

XX OS Staphylococcus aureus.

XX PN WO200170955-A2.

XX PD 27-SEP-2001.

XX XX 21-MAR-2001; 2001WO-US09180.

XX XX 21-MAR-2000; 2000US-191078P.

PR 23-MAY-2000; 2000US-206848P.

PR 26-MAY-2000; 2000US-207727P.

PR 23-OCT-2000; 2000US-242578P.

PR 27-NOV-2000; 2000US-253625P.



FT Domain 227..234 /note= "Ser/Thr kinase motif"  
FT Domain 346..364 /note= "regulatory domain"  
FT  
PN EP978562-A1.  
XX  
XX 09-FEB-2000.  
XX  
XX 06-AUG-1998; 98EP-0202654.  
XX  
XX 06-AUG-1998; 98EP-0202654.  
XX  
PA (MEDI-) DIV MEDICAL PHARMACOLOGY LEIDEN AMSTERDA.  
XX  
XX WPI: 2000-138770/13.  
XX  
XX Mammalian calcium/calmodulin dependent protein kinase related protein  
PT useful as a modulator of calcium/calmodulin dependent protein kinase  
PT activity, useful for treating seizures and damage to the central  
PT nervous system -  
XX  
PS Disclosure; Page 34-36; 48pp; English.  
XX  
XX The present sequence represents a calcium/calmodulin dependent protein  
CC kinase (CaMK-IV). The specification also describes another CaMK,  
CC designated CaMK-VI, and a CaMK-VI related peptide, designated CARP. The  
CC CARP peptide is produced by alternative splicing of the CaMK-VI gene,  
CC and plays a role in kinase-mediated neuronal plasticity which is  
CC associated with epilepsy. A large part of CARP (48 out of 55 amino  
CC acids) is highly homologous with the carboxy-terminus of human  
CC doublecortin. It is possible the calcium overload in a cell, triggered  
CC by kainate-induced seizures, leads to activation of CaMK-VI. This  
CC enhanced kinase activity induces CARP expression, which subsequently  
CC acts as a modulator of the kinase, e.g. by acting as a substrate or  
CC by irreversible binding to the activated kinase. The mammalian CARP  
CC protein may be used for the treatment and/or prevention of seizures  
CC (e.g. in epilepsy) and/or damage to the central nervous system,  
CC particularly the hippocampus. CARP peptides may also be used to treat  
CC acute brain insults (e.g. resulting from stroke, trauma, hypoxia, and  
CC ischemia), and neurodegenerative diseases (e.g. Alzheimer's disease).  
XX  
SQ Sequence 502 AA;  
Query Match 34.9%; Score 45; DB 21; Length 502;  
Best Local Similarity 58.8%; Pred. No. 2.9e+02;  
Matches 10; Conservative 3; Mismatches 4; Indels 0; Gaps 0;  
QY 2 DALEALADQTDALQSEE 18  
: : : : :  
Db 406 EAAKAAADETMKLQSEE 422  
RESULT 27  
AAW99481  
ID AAW99481 standard; Protein; 1362 AA.  
XX  
XX AAW99481;  
XX  
XX 08-JUN-1999 (first entry)  
XX  
XX Murine pCIP protein.  
XX  
XX Mouse; pCIP; p300/CBP/co-Integrator-associated protein; gene expression;  
KW breast cancer; inflammatory disease; atherosclerosis; osteoporosis; ds.  
XX  
XX Mus sp.  
XX  
XX Key Location/Qualifiers  
PH  
FT Misc-difference 540 /note= "defined in specification as ?"  
FT Misc-difference 1080 /note= "defined in specification as O"  
FT

FT Misc-difference 44..45 /note= "there appears to be 20 amino acids missing  
FT between these positions in the sequence given  
FT in the specification"  
FT  
FT Misc-difference 1044..1045 /note= "there appears to be 20 amino acids missing  
FT between these positions in the sequence given  
FT in the specification"  
XX  
PN W09856806-A1.  
XX  
XX 17-DEC-1998.  
XX  
XX 12-JUN-1998; 98WO-US12263.  
XX  
XX 12-JUN-1997; 97US-0049452.  
XX  
XX (REGC ) UNIV CALIFORNIA.  
XX  
XX Glass CK, Rose DW, Rosenfield MG, Torchia J;  
PI  
XX WPI: 1999-080883/07.  
DR N-PSDB; AAX26000.  
XX  
XX New nucleic acids encoding pCIP and NcoA-2 polypeptides - are used  
PT to identify agents that regulate gene expression, e.g. for treatment  
PT of cancer, inflammatory disease and osteoporosis  
XX  
XX Claim 16; Fig 1; 100pp; English.  
XX  
XX This sequence represents the amino acid sequence of the mouse pCIP  
CC (p300/CBP/co-integrator-associated protein) polypeptide. The protein  
CC can regulate gene expression so are potentially useful therapeutically,  
CC e.g. against (breast) cancer, inflammatory disease (e.g. atherosclerosis)  
CC or osteoporosis.  
XX  
SQ Sequence 1362 AA;  
Query Match 34.9%; Score 45; DB 20; Length 1362;  
Best Local Similarity 41.7%; Pred. No. 9.1e+02;  
Matches 10; Conservative 4; Mismatches 10; Indels 0; Gaps 0;  
QY 5 EALADQTDALQSEEAAYVKADNAA 28  
: : : : :  
Db 1053 QALSKODVFQGEAAVMDQKAA 1076  
RESULT 28  
AAW81029  
ID AAW81029 standard; Protein; 1447 AA.  
XX  
XX AAW81029;  
XX  
XX 12-MAY-1999 (first entry)  
XX  
XX Murine pCIP protein (ortholog of human AIB1 protein).  
XX  
XX AIB1; amplified in breast cancer; cancer; steroid; receptor;  
KW coactivator; SCR; estrogen; ER; estrogen dependent transcription;  
KW breast cancer; lung cancer; colon cancer; prostate cancer;  
KW melanoma.  
XX  
XX Mus musculus.  
XX  
XX Key Location/Qualifiers  
PH  
FT Region 1402  
FT /label= Potential\_C\_terminus  
FT /note= "this position encoded by stop codon"  
XX  
XX W09857982-A2.  
XX  
XX 23-DEC-1998.  
XX



CC nucleotide sequence selected from AAV43552 to AAV43560 and nucleotide  
 CC sequences identical to a known splice branch site in a known gene,  
 CC sequences identical in length to a known spacer region between splice  
 CC branch and acceptor sites in a known gene, sequences identical to a known  
 CC splice acceptor site in a known gene, sequence identical to a known  
 CC splice donor site in a known gene, an open reading frame (ORF) 3N-1  
 CC nucleotides in length, the ORF encoding a known peptide tag recognisable  
 CC by a known reaction characteristic of the known peptide tag and sequences  
 CC selected from CAGG and TAGG. The DNA sequence is inserted into the intron  
 CC within the gene to create a tagged gene, and the tagged gene is incubated  
 CC within a cell so as to maintain intact or to introduce the tagged gene  
 CC within the genome of the cell. The method is used for isolating proteins,  
 CC RNA and genes, for analysis of subcellular structures, cellular responses  
 CC and transcriptional regulation, for the study of viral infection and for  
 CC diagnosis of disease.

XX Sequence 515 AA;

Query Match 34.5%; Score 44.5; DB 19; Length 515;  
 Best Local Similarity 51.7%; Pred. No. 3.5e+02;  
 Matches 15; Conservative 4; Mismatches 7; Indels 3; Gaps 2;

QY 1 SDALEALADQTDALQSEEAAYVK--ADNA 27  
 :||:||||| : | :||| ||| | ||| |  
 Db 368 ADAVEAAAEELTA-QAEAAAKWEADKA 395

RESULT 31  
 AAW70356

ID AAW70356 standard; Protein; 516 AA.

AC AAW70356;

DT 14-DEC-1998 (first entry)

XX Chlamydomonas protein RSP3.

XX Target nucleotide; epitope; gene tagging; epitope tagging.

OS Chlamydomonas sp.

XX WO9826094-A1.

XX 18-JUN-1998.

XX 09-DEC-1997; 97WO-US22472.

XX 09-DEC-1996; 96US-0762106.

XX (JARV/) JARVIK J W.

XX Jarvik JW;

XX WPI; 1998-348546/30.

XX N-PSDB; AAV33277.

XX Epitope tagging of genes, transcripts and polypeptides - using an  
 PT oligonucleotide comprising a nucleotide sequence encoding the  
 PT epitope independently of the reading frame of the nucleotide  
 PT sequence

XX Disclosure; Fig 2A; 46pp; English.

XX The Chlamydomonas protein RSP3 is an example of a target nucleotide  
 CC sequence into which is inserted an oligonucleotide encoding an epitope.  
 CC The epitope is encoded by the oligonucleotides nucleotide sequence  
 CC independently of the reading frame. The epitope and methods disclosed  
 CC can be used for the tagging of genes, transcripts and polypeptides. They  
 CC can be used for e.g. discovering new genes, determining the size and  
 CC abundance of proteins produced by newly discovered genes, tracking the  
 CC movement of proteins within cell membranes, monitoring receptor binding  
 CC and internalisation of exogenous proteins, identifying the components of  
 CC functional protein complexes, purifying proteins, discovering the

CC function of proteins. The products, methods and uses are particularly  
 CC used for proteins that are unstable, are difficult to purify, or share  
 CC epitopes with a number of other proteins. The methods overcome the  
 CC inefficiency of epitope tagging caused by reading frame obstacles and  
 CC orientation obstacles.

XX Sequence 516 AA;

Query Match 34.5%; Score 44.5; DB 19; Length 516;  
 Best Local Similarity 51.7%; Pred. No. 3.5e+02;  
 Matches 15; Conservative 4; Mismatches 7; Indels 3; Gaps 2;

QY 1 SDALEALADQTDALQSEEAAYVK--ADNA 27  
 :||:||||| : | :||| ||| | ||| |  
 Db 369 ADAVEAAAEELTA-QAEAAAKWEADKA 396

RESULT 32

AAW70357

ID AAW70357 standard; Protein; 527 AA.

AC AAW70357;

DT 14-DEC-1998 (first entry)

XX Tagged Chlamydomonas protein RSP3.

XX Target nucleotide; epitope; gene tagging; epitope tagging.

OS Chlamydomonas sp.

XX Key Location/Qualifiers

XX Region 147-157

XX /note="Tag"

XX WO9826094-A1.

XX 18-JUN-1998.

XX 09-DEC-1997; 97WO-US22472.

XX 09-DEC-1996; 96US-0762106.

XX (JARV/) JARVIK J W.

XX Jarvik JW;

XX WPI; 1998-348546/30.

XX N-PSDB; AAV33278.

XX Epitope tagging of genes, transcripts and polypeptides - using an  
 PT oligonucleotide comprising a nucleotide sequence encoding the  
 PT epitope independently of the reading frame of the nucleotide  
 PT sequence

XX Disclosure; Fig 2B; 46pp; English.

XX The tagged Chlamydomonas protein RSP3 is an example of a target  
 CC nucleotide with an oligonucleotide encoding an epitope inserted. The  
 CC epitope is encoded by the oligonucleotides nucleotide sequence  
 CC independently of the reading frame. The epitope and methods disclosed  
 CC can be used for the tagging of genes, transcripts and polypeptides. They  
 CC can be used for e.g. discovering new genes, determining the size and  
 CC abundance of proteins produced by newly discovered genes, tracking the  
 CC movement of proteins within cell membranes, monitoring receptor binding  
 CC and internalisation of exogenous proteins, identifying the components of  
 CC functional protein complexes, purifying proteins, discovering the  
 CC function of proteins. The products, methods and uses are particularly  
 CC used for proteins that are unstable, are difficult to purify, or share  
 CC epitopes with a number of other proteins. The methods overcome the  
 CC inefficiency of epitope tagging caused by reading frame obstacles and  
 CC orientation obstacles.

SQ Sequence 527 AA;

Query Match 34.5%; Score 44.5; DB 19; Length 527;  
Best Local Similarity 51.7%; Pred. No. 3.6e+02;  
Matches 15; Conservative 4; Mismatches 7; Indels 3; Gaps 2;

Qy 1 SDALFALADQTDALQSEEAAYVK--ADNA 27  
:||:| | : | :||| | | |  
Db 380 ADAVEAAALTA-QAEAAAKWEADKA 407

RESULT 33  
AAW61908  
ID AAW61908 standard; Protein; 552 AA.  
XX  
AC AAW61908;  
XX  
DT 16-SEP-1998 (first entry)  
XX  
DE Tandem tagged RSP3 pf14 gene product) protein.  
XX  
KW Tagged gene; tagged transcript; hybrid intron; protein tag;  
KW protein isolation; recombination; subcellular structure analysis;  
KW transcriptional regulation; viral infection; Chlamydomonas; pf14 gene.  
XX  
OS Chlamydomonas sp.  
OS Synthetic.  
XX  
PN WO9820031-A1.  
XX  
PD 14-MAY-1998.  
XX  
PF 07-NOV-1997; 97WO-US20150.  
XX  
PR 08-NOV-1996; 96US-0705404.  
XX  
PA (JARV/) JARVIK J W.  
XX  
PI Jarvik JW;  
XX  
DR WPT; 1998-286861/25.  
XX  
PT Tagging genes, transcripts and proteins - using tag-creating DNA  
PT inserted into intron of gene to create 2 hybrid introns separated by  
PT new exon encoding protein tag  
XX  
PS Disclosure; Pages 22-23; 56pp; English.

XX This sequence represents a tandem tagged RSP3 protein, a Chlamydomonas  
CC pf14 gene product. The gene can be tagged by using the method of  
CC invention of tagging genes, transcripts and proteins in cells in a  
CC single recombinational event. The method comprises producing a tagged  
CC gene by inserting a DNA sequence into an intron of a gene by selecting a  
CC DNA sequence having a 5' portion free of any nucleotide sequence  
CC selected from AAV43548 to AAV43551, a nucleotide sequence selected from  
CC AAV43552 to AAV43560 and nucleotide sequences identical to a known splice  
CC branch site in a known gene, sequences identical in length to a known  
CC spacer region between splice branch and acceptor sites in a known gene,  
CC sequences identical to a known splice acceptor site in a known gene,  
CC sequence identical to a known splice donor site in a known gene, an open  
CC reading frame (ORF) 3N-1 nucleotides in length, the ORF encoding a known  
CC peptide tag recognisable by a known reaction characteristic of the known  
CC peptide tag and sequences selected from CAGG and TAGG. The DNA sequence  
CC is inserted into the intron within the gene to create a tagged gene, and  
CC the tagged gene is incubated within a cell so as to maintain intact or  
CC to introduce the tagged gene within the genome of the cell. The method  
CC is used for isolating proteins, RNA and genes, for analysis of  
CC subcellular structures, cellular responses and transcriptional  
CC regulation, for the study of viral infection and for diagnosis of  
CC disease.

XX  
SQ Sequence 552 AA;

Query Match 34.5%; Score 44.5; DB 19; Length 552;  
Best Local Similarity 51.7%; Pred. No. 3.8e+02;  
Matches 15; Conservative 4; Mismatches 7; Indels 3; Gaps 2;

Qy 1 SDALFALADQTDALQSEEAAYVK--ADNA 27  
:||:| | : | :||| | | |  
Db 405 ADAVEAAALTA-QAEAAAKWEADKA 432

RESULT 34  
ABG01077  
ID ABG01077 standard; Protein; 42 AA.  
XX  
AC ABG01077;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #1068.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DR WPI; 2001-639362/73.  
DR N-PSDB; AAS65264.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX  
PS Claim 20; SEQ ID NO 31436; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX  
SQ Sequence 42 AA;

Query Match 34.1%; Score 44; DB 22; Length 42;  
Best Local Similarity 36.4%; Pred. No. 22;





PR	13-AUG-1999;	99US-0148565.	RESULT 36
PR	13-AUG-1999;	99US-0148684.	AAG61028
PR	16-AUG-1999;	99US-0149368.	ID AAG61028 standard; Protein; 274 AA.
PR	17-AUG-1999;	99US-0149175.	XX
PR	18-AUG-1999;	99US-0149426.	AC AAG61028;
PR	20-AUG-1999;	99US-0149722.	XX
PR	20-AUG-1999;	99US-0149723.	DT 18-OCT-2000 (first entry)
PR	20-AUG-1999;	99US-0149929.	XX
PR	23-AUG-1999;	99US-0149902.	XX
PR	23-AUG-1999;	99US-0149930.	DE Arabidopsis thaliana protein fragment SEQ ID NO: 79109.
PR	25-AUG-1999;	99US-0150566.	XX
PR	26-AUG-1999;	99US-0150884.	KW Protein identification; signal transduction pathway; metabolic pathway;
PR	27-AUG-1999;	99US-0151065.	KW hybridisation assay; genetic mapping; gene expression control; promoter;
PR	27-AUG-1999;	99US-0151066.	XX termination sequence.
PR	27-AUG-1999;	99US-0151080.	XX
PR	30-AUG-1999;	99US-0151303.	OS Arabidopsis thaliana.
PR	31-AUG-1999;	99US-0151438.	XX
PR	01-SEP-1999;	99US-0151930.	PN EP1033405-A2.
PR	07-SEP-1999;	99US-0152363.	XX
PR	10-SEP-1999;	99US-0153070.	PD 06-SEP-2000.
PR	13-SEP-1999;	99US-0153758.	XX
PR	15-SEP-1999;	99US-0154018.	PF 25-FEB-2000; 2000EP-0301439.
PR	16-SEP-1999;	99US-0154039.	XX
PR	20-SEP-1999;	99US-0154779.	PR 25-FEB-1999; 99US-0121825.
PR	22-SEP-1999;	99US-0155139.	PR 05-MAR-1999; 99US-0123180.
PR	23-SEP-1999;	99US-0155486.	PR 09-MAR-1999; 99US-0123548.
PR	24-SEP-1999;	99US-0155659.	PR 23-MAR-1999; 99US-0125788.
PR	28-SEP-1999;	99US-0156458.	PR 25-MAR-1999; 99US-0126264.
PR	29-SEP-1999;	99US-0156596.	PR 29-MAR-1999; 99US-0126785.
PR	04-OCT-1999;	99US-0157117.	PR 01-APR-1999; 99US-0127462.
PR	05-OCT-1999;	99US-0157753.	PR 06-APR-1999; 99US-0128234.
PR	06-OCT-1999;	99US-0157865.	PR 08-APR-1999; 99US-0128714.
PR	07-OCT-1999;	99US-0158029.	PR 16-APR-1999; 99US-0129845.
PR	08-OCT-1999;	99US-0158232.	PR 19-APR-1999; 99US-0130077.
PR	12-OCT-1999;	99US-0158369.	PR 21-APR-1999; 99US-0130449.
PR	13-OCT-1999;	99US-0159293.	PR 23-APR-1999; 99US-0130510.
PR	13-OCT-1999;	99US-0159294.	PR 28-APR-1999; 99US-0130891.
PR	13-OCT-1999;	99US-0159295.	PR 30-APR-1999; 99US-0131449.
PR	14-OCT-1999;	99US-0159329.	PR 30-APR-1999; 99US-0132048.
PR	14-OCT-1999;	99US-0159330.	PR 04-MAY-1999; 99US-0132407.
PR	14-OCT-1999;	99US-0159331.	PR 05-MAY-1999; 99US-0132484.
PR	14-OCT-1999;	99US-0159637.	PR 06-MAY-1999; 99US-0132485.
PR	14-OCT-1999;	99US-0159638.	PR 06-MAY-1999; 99US-0132486.
PR	18-OCT-1999;	99US-0159584.	PR 07-MAY-1999; 99US-0132487.
PR	21-OCT-1999;	99US-0160741.	PR 07-MAY-1999; 99US-0132863.
PR	21-OCT-1999;	99US-0160767.	PR 11-MAY-1999; 99US-0134256.
PR	21-OCT-1999;	99US-0160768.	PR 14-MAY-1999; 99US-0134218.
PR	21-OCT-1999;	99US-0160770.	PR 14-MAY-1999; 99US-0134219.
PR	21-OCT-1999;	99US-0160814.	PR 14-MAY-1999; 99US-0134221.
PR	21-OCT-1999;	99US-0160815.	PR 18-MAY-1999; 99US-0134768.
PR	22-OCT-1999;	99US-0160980.	PR 19-MAY-1999; 99US-0134941.
PR	22-OCT-1999;	99US-0160981.	PR 20-MAY-1999; 99US-0135124.
PR	22-OCT-1999;	99US-0160989.	PR 21-MAY-1999; 99US-0135353.
PR	25-OCT-1999;	99US-0161404.	PR 24-MAY-1999; 99US-0135629.
PR	25-OCT-1999;	99US-0161405.	PR 25-MAY-1999; 99US-0136021.
PR	25-OCT-1999;	99US-0161406.	PR 27-MAY-1999; 99US-0136392.
PR	26-OCT-1999;	99US-0161359.	PR 28-MAY-1999; 99US-0136782.
PR	26-OCT-1999;	99US-0161360.	PR 01-JUN-1999; 99US-0137222.
PR	26-OCT-1999;	99US-0161361.	PR 03-JUN-1999; 99US-0137528.
PR	28-OCT-1999;	99US-0161920.	PR 04-JUN-1999; 99US-0137502.
PR	28-OCT-1999;	99US-0161992.	PR 07-JUN-1999; 99US-0137724.
PR	28-OCT-1999;	99US-0161993.	PR 08-JUN-1999; 99US-0138094.
PR	29-OCT-1999;	99US-0162142.	PR 10-JUN-1999; 99US-0138540.
PR			PR 10-JUN-1999; 99US-0138847.
PR			PR 14-JUN-1999; 99US-0139119.
PR			PR 16-JUN-1999; 99US-0139452.
PR			PR 16-JUN-1999; 99US-0139453.
PR			PR 17-JUN-1999; 99US-0139492.
PR			PR 18-JUN-1999; 99US-0139454.
PR			PR 18-JUN-1999; 99US-0139455.
PR			PR 18-JUN-1999; 99US-0139456.
PR			PR 18-JUN-1999; 99US-0139457.
PR			PR 18-JUN-1999; 99US-0139458.

Query Match 34.18; Score 44; DB 21; Length 274;

Best Local Similarity 39.18; Pred. No. 2e+02;

Matches 9; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

Oy 5 EALADOTDALQSEEAAYVKADNA 27

||| : || : || : || : |||

Db 53 EALQORDKALSERDKALIERDNA 75

PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.

PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 34.1%; Score 44; DB 21; Length 274;  
Best Local Similarity 39.1%; Pred. No. 2e+02;  
Matches 9; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

QY 5 EALADOTDALQSEAAVVKADNA 27  
||| : || : ||| : |||  
Db 53 EALQQRDKALSERDKALIERDNA 75

RESULT 37  
AAM78512  
ID AAM78512 standard; Protein; 284 AA.  
XX  
AC AAM78512;  
XX  
DT 06-NOV-2001 (first entry)

XX	Human protein SEQ ID NO 1174.	DE	Arabidopsis thaliana protein fragment SEQ ID NO: 75230.
XX		XX	
KW	Human; cytokine; cell proliferation; cell differentiation; gene therapy;	KW	Protein identification; signal transduction pathway; metabolic pathway;
KW	vaccine; peptide therapy; stem cell growth factor; haematopoiesis;	KW	hybridisation assay; genetic mapping; gene expression control; promoter;
KW	tissue growth factor; immunomodulatory; cancer; leukaemia;	XX	termination sequence.
KW	nervous system disorder; arthritis; inflammation.	OS	Arabidopsis thaliana.
XX		XX	
OS	Homo sapiens.	PN	EP1033405-A2.
XX		XX	
PN	WO200157190-A2.	PD	06-SEP-2000.
XX		XX	
PD	09-AUG-2001.	PF	25-FEB-2000; 2000EP-0301439.
XX		XX	
PF	05-FEB-2001; 2001WO-US04098.	PR	25-FEB-1999; 99US-0121825.
XX		PR	05-MAR-1999; 99US-0123180.
XX	03-FEB-2000; 2000US-0496914.	PR	09-MAR-1999; 99US-0123548.
PR	27-APR-2000; 2000US-0560875.	PR	23-MAR-1999; 99US-0125788.
PR	20-JUN-2000; 2000US-0598075.	PR	29-MAR-1999; 99US-0126264.
PR	19-JUL-2000; 2000US-0620325.	PR	29-MAR-1999; 99US-0126785.
PR	01-SEP-2000; 2000US-0654936.	PR	01-APR-1999; 99US-0127462.
PR	15-SEP-2000; 2000US-0663561.	PR	06-APR-1999; 99US-0128234.
PR	20-OCT-2000; 2000US-0693325.	PR	08-APR-1999; 99US-0128714.
PR	30-NOV-2000; 2000US-0728422.	PR	16-APR-1999; 99US-0129845.
XX		PR	19-APR-1999; 99US-0130077.
PA	(HYSE-) HYSEQ INC.	PR	21-APR-1999; 99US-0130449.
XX		PR	23-APR-1999; 99US-0130510.
PI	Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;	PR	23-APR-1999; 99US-0130891.
PI	Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;	PR	28-APR-1999; 99US-0131449.
PI	Xue AJ, Yang Y, . Wejhrman T, Goodrich R;	PR	30-APR-1999; 99US-0132048.
XX		PR	30-APR-1999; 99US-0132407.
DR	WPI; 2001-476283/51.	PR	04-MAY-1999; 99US-0132484.
DR	N-PSDB; AAK51645.	PR	05-MAY-1999; 99US-0132485.
XX		PR	06-MAY-1999; 99US-0132486.
PT	Nucleic acids encoding polypeptides with cytokine-like activities,	PR	06-MAY-1999; 99US-0132487.
PT	useful in diagnosis and gene therapy -	PR	07-MAY-1999; 99US-0132863.
XX		PR	11-MAY-1999; 99US-0134256.
PS	Claim 20; Page 3417-3418; 6221pp; English.	PR	14-MAY-1999; 99US-0134218.
XX		PR	14-MAY-1999; 99US-0134219.
CC	The invention relates to polynucleotides (AAK51456-AAK53435) and the	PR	14-MAY-1999; 99US-0134221.
CC	encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to	PR	14-MAY-1999; 99US-0134370.
CC	cytokine, cell proliferation or cell differentiation or which may induce	PR	18-MAY-1999; 99US-0134766.
CC	production of other cytokines in other cell populations. The	PR	19-MAY-1999; 99US-0134941.
CC	polynucleotides and polypeptides are useful in gene therapy, vaccines or	PR	20-MAY-1999; 99US-0135124.
CC	peptide therapy. The polypeptides have various cytokine-like activities,	PR	21-MAY-1999; 99US-0135353.
CC	e.g. stem cell growth factor activity, haematopoiesis regulating	PR	24-MAY-1999; 99US-0135629.
CC	activity, tissue growth factor activity, immunomodulatory activity and	PR	25-MAY-1999; 99US-0136021.
CC	activin/inhibin activity and may be useful in the diagnosis and/or	PR	27-MAY-1999; 99US-0136392.
CC	treatment of cancer, leukaemia, nervous system disorders, arthritis and	PR	28-MAY-1999; 99US-0136782.
CC	inflammation.	PR	01-JUN-1999; 99US-0137222.
CC	Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666	PR	03-JUN-1999; 99US-0137528.
CC	(AAM80020) are omitted as the relevant pages from the sequence listing	PR	04-JUN-1999; 99US-0137502.
CC	were missing at the time of publication.	PR	07-JUN-1999; 99US-0137724.
XX		PR	08-JUN-1999; 99US-0138094.
SQ	Sequence 284 AA;	PR	10-JUN-1999; 99US-0138540.
		PR	10-JUN-1999; 99US-0138847.
	Query Match 34.1%; Score 44; DB 22; Length 284;	PR	14-JUN-1999; 99US-0139119.
	Best Local Similarity 47.6%; Pred. No. 2e+02;	PR	16-JUN-1999; 99US-0139452.
	Matches 10; Conservative 3; Mismatches 8; Indels 0; Gaps 0;	PR	16-JUN-1999; 99US-0139453.
QY	5 EALADOTDALQSEEAAYVKAD 25	PR	17-JUN-1999; 99US-0139492.
	i:	PR	18-JUN-1999; 99US-0139454.
Db	62 EALKDAEDSLAAEEAAAKAE 82	PR	18-JUN-1999; 99US-0139455.
		PR	18-JUN-1999; 99US-0139456.
RESULT 38		PR	18-JUN-1999; 99US-0139457.
AGS58292		PR	18-JUN-1999; 99US-0139458.
ID	AGS58292 standard; Protein; 296 AA.	PR	18-JUN-1999; 99US-0139459.
XX		PR	18-JUN-1999; 99US-0139460.
AC	AGS58292;	PR	18-JUN-1999; 99US-0139461.
XX		PR	18-JUN-1999; 99US-0139462.
XX		PR	18-JUN-1999; 99US-0139463.
DT	18-OCT-2000 (first entry)	PR	18-JUN-1999; 99US-0139750.
XX		PR	18-JUN-1999; 99US-0139763.
		PR	21-JUN-1999; 99US-0139817.

PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 08-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.

PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 18-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161922.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 34.1%; Score 44; DB 21; Length 296;  
Best Local Similarity 39.1%; Pred. No. 2.2e+02;

Matches 9; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

Qy 5 EALADQTDALQSEEAAYVKADNA 27  
||| : || : ||| : |||  
Db 75 EALQQRDKALSERDKALIERDNA 97

RESULT 39

AAG61027

ID AAG61027 standard; Protein; 296 AA.

XX AAG61027;

XX 18-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 79108.

XX Protein identification; signal transduction pathway; metabolic pathway;  
XX hybridisation assay; genetic mapping; gene expression control; promoter;  
XX termination sequence.

OS Arabidopsis thaliana.



PR	16-SEP-1999;	99US-01540399.
PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.
PR	14-OCT-1999;	99US-0159325.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	23-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.

```

Query Match      34.1%; Score 44; DB 21; Length 296;
Best Local Similarity 39.1%; Pred. No. 2.2e+02;
Matches 9; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

QY 5 EALADQTDALQSBEEAAVVKADNA 27
    |||  ||  |||  |||  |||  |||
Db 75 EALQORDKALSERDKALIERDNA 97

```

RESULT 40  
AAW61234  
ID AAW61234 standard: Protein: 344 AA.

AC	AAW61234;	
XX		
DT	02-OCT-1998 (first entry)	
XX		
DE	Streptococcus pneumoniae	
XX	Spl09 protein.	
KW	Streptococcus pneumoniae;	
XX	antigen; vaccine; infection;	
KW	detection; pneumonia;	
XX	otitis media; meningitis.	
OS	Streptococcus pneumoniae.	
XX		
PN	WO9818930-A2.	
XX		
PD	07-MAY-1998.	
XX		
PF	30-OCT-1997;	
XX	97WO-US19422.	
XX		
PR	31-OCT-1996;	
XX	96US-0029960.	

XX	(HUMA-) HUMAN GENOME SCI INC.
XX	
XX	Choi GH, Hromockyj A, Johnson LS, Kunsch CA;
XX	
DR	WPI: 1998-272224/24.
DR	N-PSDB: RAV27419.
XX	
PT	Nucleic acid encoding antigenic peptide(s) from Streptococcus
PT	pneumoniae - or their epitope-containing fragments, useful in
PT	protective or therapeutic vaccines, and for diagnosis
XX	
XX	Claim 11; Page 87; 118pp; English.
XX	
CC	The present sequence represents a protein from Streptococcus
CC	The nucleic acid sequence encoding the Streptococcus pneumoniae
CC	can be useful in vaccines for inducing protective antibodies
CC	Streptococcus pneumoniae, for treatment or prevention of infection
CC	pneumonia, otitis media or meningitis. Probes based on the nu-
CC	are used to detect Streptococcus infection (by usual hybridization
CC	amplification methods), also for isolating Streptococcus genes
CC	allelic variants. The protein can be used similarly to detect
CC	antibodies in standard immunoassays, especially for diagnosing
CC	monitoring infections. Antibodies which bind the protein are
CC	detect corresponding antigens, to purify the protein and for
CC	immunisation (optionally coupled to a toxin). Vaccines are ad-
CC	e.g. by injection, orally or through the skin, typically at 0
CC	(especially 10-300) mu g/ml per dose.
XX	
SQ	Sequence 344 AA;
	Query Match 34.1%; Score 44; DB 19; Length 344;
	Best Local Similarity 55.0%; Pred. NO. 2.6e+02;
	Matches 11; Conservative 2; Mismatches 7; Indels 0
QY	8 ADOTDALQSEEAUVKADNA 27          :   :   :   :
Dd	3 AGOTDASQIEKAARVSGGKA 22
	Search completed: October 13, 2002, 04:45:34
	Job time : 12.3797 secs

THIS PAGE BLANK (USPTO)



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 13, 2002, 04:50:13 ; Search time 1877.39 Seconds  
(without alignments)  
1143.085 Million cell updates/sec

Title: US-09-847-539A-6

Perfect score: 777

Sequence: 1 VDSPIEQRIIPNGTTLNL.....KKEERQNVNLTPTTGESNP 159

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters: -DEV=xlh

-Q/cgn2\_1/USPTO\_spool/US09847539/runat\_10102002\_093106\_5005/app\_query.fasta\_1.526  
-DB=EST -OPMT=fastap -SUFFIX=std.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45  
-DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=40 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09847539 @CGN\_1\_1\_763 @runat\_10102002\_093106\_5005 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAP -THREADES=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-WARN\_TIMEOUT=30 -LARGE=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-XGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_htc:\*  
12: gb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_inv:\*  
15: em\_gss\_pln:\*  
16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	127.5	16.4	938	12 CNS07CG2	AL439128 T7 end of
2	112.5	14.5	689	10 BF480699	BF480699 L0-2660T3

c 3	112.5	14.5	1061	12 CNS0738M	AL427196 clone BA0
c 4	112	14.4	442	9 AU180004	AU180004 AU180004
c 5	112	14.4	475	10 BM029212	BM029212 IpSkn0005
c 6	112	14.4	570	9 AU179963	AU179963 AU179963
c 7	111.5	14.4	615	9 AU105781	AU105781 SWOVL3CAN
c 8	111	14.3	344	9 AU180137	AU180137 AU180137
c 9	110.5	14.2	617	10 BM355271	BM355271 rr22h07.y
c 10	110	14.2	517	9 AU179732	AU179732 AU179732
c 11	110	14.2	610	9 AU180091	AU180091 AU180091
c 12	110	14.2	610	9 AU180168	AU180168 AU180168
c 13	109.5	14.1	453	9 AU179269	AU179269 AU179269
c 14	108.5	14.0	715	10 BJ064193	BJ064193 BJ064193
c 15	108	13.9	548	9 AU177381	AU177381 AU177381
c 16	107	13.8	380	9 AU179899	AU179899 AU179899
c 17	107	13.8	664	10 BG443972	BG443972 GA_Ea002
c 18	107	13.8	857	10 BM360707	BM360707 GA_Ea003
c 19	107	13.8	881	10 BG445649	BG445649 GA_Ea002
c 20	107	13.8	887	10 BG444821	BG444821 GA_Ea002
c 21	106	13.6	546	9 AI959039	AI959039 fd06g08.y
c 22	106	13.6	620	10 BJ114733	BJ114733 BJ114733
c 23	106	13.6	947	12 CNS0794Y	AL434840 T3 end of
c 24	105.5	13.6	653	10 BJ001775	BJ001775 BJ001775
c 25	105.5	13.6	669	10 BJ000182	BJ000182 BJ000182
c 26	105.5	13.6	687	10 BJ011593	BJ011593 BJ011593
c 27	105.5	13.6	710	10 BJ001283	BJ001283 BJ001283
c 28	105.5	13.6	968	12 CNS07CXJ	AL439757 T7 end of
c 29	105	13.5	473	10 BE405838	BE405838 WHE0437.A
c 30	105	13.5	755	9 AJ395963	AJ395963 AJ395963
c 31	104	13.4	530	10 BM300555	BM300555 MCA053H02
c 32	103.5	13.3	616	10 BJ006142	BJ006142 BJ006142
c 33	103.5	13.3	666	10 BJ006464	BJ006464 BJ006464
c 34	103	13.3	938	12 A2676392	A2676392 ENTHV47TR
c 35	102.5	13.2	537	10 BG930118	BG930118 etshEST05
c 36	102.5	13.2	628	12 A0648620	AQ48620 RPC193-BC
c 37	102.5	13.2	790	12 CNS07270	AL425866 clone BA0
c 38	102	13.1	727	10 BG440614	BG440614 GA_Ea000
c 39	102	13.1	1076	12 CNS06MOG	AL405734 T3 end of
c 40	101.5	13.1	868	12 BH161951	BH161951 ENTRB12TR
c 41	101	13.0	609	10 BJ119410	BJ119410 BJ119410
c 42	101	13.0	619	9 AU180092	AU180092 AU180092
c 43	101	13.0	936	12 CNS07CJY	AL439288 T7 end of
c 44	100.5	12.9	864	12 A2547838	A2547838 ENTF21TR
c 45	99.5	12.8	522	10 B1215589	B1215589 RE22139.5

## ALIGNMENTS

RESULT 1	CNS07CG2	938 bp	DNA	linear	GSS 08-JUL-2001
CNS07CG2/c	T7 end of clone BD0AA002F07 of library BD0AA from strain CBS 94 of				
LOCUS	Candida tropicalis, genomic survey sequence.				
DEFINITION	AL439128				
ACCESSION	AL439128.1	GI:12222541			
VERSION	GSS.				
KEYWORDS	Candida tropicalis.				
SOURCE	Candida tropicalis.				
ORGANISM	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; mitosporic Saccharomycetales; Candida.				
REFERENCE	1 (bases 1 to 938)				
AUTHORS	Blandin, G., Ozler-Kalogeropoulos, O., Wincker, P., Artiguenave, F. and Dujon, B.				
TITLE	Genomic exploration of the hemiascomycetous yeasts: 16. Candida tropicalis				
JOURNAL	FEBS Lett. 487 (1), 91-94 (2000)				
MEDLINE	20584726				
REFERENCE	2 (bases 1 to 938)				
AUTHORS	Souci, J.L., Aigle, M., Artiguenave, F., Blandin, G., Bolotin-Fukuhara, M., Bon, E., Brottier, P., Casaregola, S., de-Montigny, J., Dujon, B., Durrens, P., Lepingle, A., Llorente, B., Malpertuy, A., Neugebuse, C., Ozler-Kalogeropoulos, O., Potier, S., Saurin, W., Tekala, F., Toffano-Nioche, C., Wesolowski-Louvel, M., Wincker, P. and Weissensbach, J.				









1997



```

Db 366 TTGAAGGATACTGAGGCTGCTGAGGCTGCTGAGGCTGCTTAGGATTCTGAGGCTGCTGAG 307
Qy 64 AlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaValValLysAla--- 82
Db 306 GCTGCTTAGGATTCTGAGGATACTGAGGCTGCTGATCATTTCTGATGTTTGAAGGCTACT 247
Qy 83 -----AspAsnAlaAlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAsp 98
Db 246 GAGGCTGTTTAGGATTCTGAGGATACTGAGGCTGATGAGGCTGCTTAGGATTCTGAGGAT 187
Qy 99 AlaLeuGlnSerGluGluAlaGluValValGlnSerAspAsnAlaAlaSerAspAlaTrp 118
Db 186 ACTGAGCGGCTGAGGATTCTGAGGTTTGAAGGCTACTGAGGCTGCTGAGGATTCTGAG 127
Qy 119 GluLysAlaAlaThrProLeuAlaLeuAspValLysLysThrLysAspThrLysProVal 138
Db 126 GATGTTGAGGCT-----GCTGAGGATTCTGAGGTTTGAAGGATACTGAGGCTGCT 76
Qy 139 ValLysLysGluGlu 143
Db 75 GAGGATTCTGAGGAT 61

RESULT 12
LOCUS AU180168 Medaka liver cDNA library (OLE) from HNI Oryzias latipes
DEFINITION cDNA clone OLE20.04g similar to pir|A48048| egg envelope protein wf
-winter flounder, mRNA sequence.
ACCESSION AU180168
VERSION AU180168.1 GI:13429005
KEYWORDS EST.
SOURCE Japanese medaka.
ORGANISM Oryzias latipes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percormorpha; Atherinomorpha;
Belontiiformes; Adrianchthyidae; Oryziinae; Oryzias.
REFERENCE 1 (bases 1 to 610)
AUTHORS Naruse,K., Mitani,H. and Tanaka,M.
TITLES Medaka EST project in University of Tokyo (2001)
JOURNAL Unpublished (2001)
COMMENT Contact: Kiyoshi Naruse
Department of Biological Sciences
Graduate School of Science, University of Tokyo
Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, Japan
Tel: 81-3-5841-4443
Fax: 81-3-5841-4410
Email: naruse@biol.s.u-tokyo.ac.jp
This clone was isolated from Medaka liver cDNA library (OLE).

FEATURES
Source
1..610
/organism="Oryzias latipes"
/strain="HNI"
/db_xref="taxon:8090"
/clone="OLE20.04g"
/clone_lib="Medaka liver cDNA library (OLE) from HNI"
/tissue_type="liver"
/dev_stage="adult"
BASE COUNT 173 a 189 c 104 g 144 t
ORIGIN

Alignment Scores:
Pred. No.: 0.0797 Length: 610
Score: 110.00 Matches: 40
Percent Similarity: 46.90% Conservative: 28
Best Local Similarity: 27.59% Mismatches: 55
Query Match: 14.16% Indels: 22
DB: 9 Gaps: 4

US-09-847-539A-6 (1-159) x AU180168 (1-610)
Qy 18 ThrAsnLeuGlyAsnAlaProGluLysLeuAlaLeuArg----- 31

```

```

Db 438 AGCAGTTTATTGTCACACTCCGAGCAGTGTGTTTGAAGAGCCACAAAGGTATTTTAT 379
Qy 32 -----AsnGluGluArgAlaIleAspGluLeuLysLysGlnAla 44
Db 378 AATPACTCATCTACTCACAAGAGTGAANTGTTGATGGTGGAGTCTGAGGATT 319
Qy 45 IleGluAspLysGluAlaThrAlaIleGluAlaAla---SerSerAspAlaLeuGlu 63
Db 318 TTGAAGGATACTGAGGCTGCTGAGGCTGCTGAGGCTCTTAGGATTCTGAGGCTGCTGAG 259
Qy 64 AlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLysAla--- 82
Db 258 GCTGCTTAGGATTCTGAGGATACTGAGGCTGCTGATCATTTCTGATGTTTGAAGGTTACT 199
Qy 83 -----AspAsnAlaAlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAsp 98
Db 198 GAGGCTGTTTAGGATTCTGAGGATACTGAGGCTGCTGAGGCTGCTTAGGATTCTGAGGAT 139
Qy 99 AlaLeuGlnSerGluGluAlaGluValValGlnSerAspAsnAlaAlaSerAspAlaTrp 118
Db 138 ACTGAGCGGCTGAGGATTCTGAGGTTTGAAGGCTACTGAGGCTGCTGAGGATTCTGAG 79
Qy 119 GluLysAlaAlaThrProLeuAlaLeuAspValLysLysThrLysAspThrLysProVal 138
Db 78 GATGTTGAGGCT-----GCTGAGGATTCTGAGGTTTGAAGGATACTGAGGCTGCT 28
Qy 139 ValLysLysGluGlu 143
Db 27 GAGGATTCTGAGGAT 13

RESULT 13
LOCUS AU179269/c
DEFINITION AU179269 Medaka liver cDNA library (OLE) from HNI Oryzias latipes
cDNA clone OLE06.03e similar to gb|AB025967| Oryzias latipes mRNA
for choriolegen in Hminor, complete cds, mRNA sequence.
ACCESSION AU179269
VERSION AU179269.1 GI:13428106
KEYWORDS EST.
SOURCE Japanese medaka.
ORGANISM Oryzias latipes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percormorpha; Atherinomorpha;
Belontiiformes; Adrianchthyidae; Oryziinae; Oryzias.
REFERENCE 1 (bases 1 to 453)
AUTHORS Naruse,K., Mitani,H. and Tanaka,M.
TITLES Medaka EST project in University of Tokyo (2001)
JOURNAL Unpublished (2001)
COMMENT Contact: Kiyoshi Naruse
Department of Biological Sciences
Graduate School of Science, University of Tokyo
Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, Japan
Tel: 81-3-5841-4443
Fax: 81-3-5841-4410
Email: naruse@biol.s.u-tokyo.ac.jp
This clone was isolated from Medaka liver cDNA library (OLE).

FEATURES
Source
1..453
/organism="Oryzias latipes"
/strain="HNI"
/db_xref="taxon:8090"
/clone="OLE06.03e"
/clone_lib="Medaka liver cDNA library (OLE) from HNI"
/tissue_type="liver"
/dev_stage="adult"
BASE COUNT 137 a 159 c 59 g 98 t
ORIGIN

Alignment Scores:
Pred. No.: 0.0631 Length: 453
Score: 109.50 Matches: 41

```







```

source
1..664
/organism="Gossypium arboreum"
/strain="AKA"
/cultivar="8400"
/db_xref="taxon:29729"
/clone="GA_Ea0022N09f"
/clone_lib="Gossypium arboreum 7-10 dpa fiber library"
/tissue_type="Fibers isolated from bolls harvested 7-10
dpa"
/lab_host="E. coli"
/note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"
BASE COUNT 221 a 124 c 177 g 141 t 1 others
ORIGIN

Alignment Scores:
Pred. No.: 0.179 Length: 664
Score: 107.00 Matches: 39
Percent Similarity: 44.97% Conservative: 28
Best Local Similarity: 26.17% Mismatches: 58
Query Match: 13.77% Indels: 24
DB: 10 Gaps: 5

US-09-847-539A-6 (1-159) x BG443972 (1-664)
QY 9 ArgilleProAsnGlyGlyThrLeuThrAsnLeuLeuGlyAsnAlaProGluLysLeu 28
Db 23 CAATTATTTCCCACTTCCTTTTCCCTCAAAATTTCTTGGGATTCCTATGGCTACTGCT 82
QY 29 AlaLeuArgAsnGluGluArgAlaIleAspGluLeuLys---LysGlnAlaIleGluAsp 47
Db 83 GAGGTTGCATCAGTTCAAACTGCACCTTCATGAGGAGAAACTGAAGATCAGTGAAGGAT 142
QY 48 LysGluAlaThrThrAlaIleGluAlaIleSerSerAspAlaLeuGluAlaLeuAlaAsp 67
Db 143 CAAGAAGCTACACATGATGAGTAGTTGCCGAAGCTTCCTGCAGCGAAGAGCTGCCGAA 202
QY 68 GlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLysAlaAspAsnAlaIleAsp 87
Db 203 GAGCCAAAG-----GAAGCGAGCGCTGGCGGCTGCAGTCGAGGAACCTGAAGCCTCG 256
QY 88 AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaGluVal 107
Db 257 -----GTTGAAGTTGAAACCAAGGAGGTG 280
QY 108 ValGlnSerAspAsnAlaAlaSerAspAlaTrpGluLysAlaAlaThrProIleAlaLeu 127
Db 281 GTAGAAGAACCAAGGCTGTACTGAG-----GAGCCAGTAGTAGAG 322
QY 128 AspValLysLysThrLysAspThrLysProVal-----ValLysLysGluGluArg 144
Db 323 GAGACTCCTAAGGAACACAGTCGAGAACCAAGCTTCGCGAGGAGATCAAGGAACCACTGAA 382
QY 145 GlnAsnValAsnThrLeuProThrThr 153
Db 383 CAAACCGTTGAAACCAAGAACTACA 409

RESULT 18
BM360707 857 bp mRNA linear EST 09-JAN-2002
LOCUS
DEFINITION
GA_Ea0034C23r Gossypium arboreum 7-10 dpa fiber library Gossypium
arboreum cDNA clone GA_Ea0034C23r, mRNA sequence.
ACCESSION
BM360707
VERSION
BM360707.1 GI:18101453
KEYWORDS
EST.
SOURCE
Gossypium arboreum.
ORGANISM
Gossypium arboreum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.
1 (bases 1 to 857)
Wing,R.A., Frisch,D., Yu,Y., Main,D., Rambo,T., Simmons,J., Henry
,D., Wood,T.C., Leslie,A. and Wilkins,T.A.
An integrated analysis of the genetics, development, and evolution

```

```

of the cotton fiber
Unpublished (2000)
Contact: Wing RA
Clemson University Genomics Institute
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Total High Quality bases = 618
Seq primer: TAATACGACTCACTATAGG
High quality sequence stop: 846.
FEATURES
Location/Qualifiers
1..857
/organism="Gossypium arboreum"
/strain="AKA"
/cultivar="8400"
/db_xref="taxon:29729"
/clone="GA_Ea0034C23r"
/clone_lib="Gossypium arboreum 7-10 dpa fiber library"
/tissue_type="Fibers isolated from bolls harvested 7-10
dpa"
/lab_host="E. coli"
/note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"
BASE COUNT 271 a 150 c 230 g 206 t
ORIGIN

Alignment Scores:
Pred. No.: 0.243 Length: 857
Score: 107.00 Matches: 39
Percent Similarity: 44.97% Conservative: 28
Best Local Similarity: 26.17% Mismatches: 58
Query Match: 13.77% Indels: 24
DB: 10 Gaps: 5

US-09-847-539A-6 (1-159) x BM360707 (1-857)
QY 9 ArgilleProAsnGlyGlyThrLeuThrAsnLeuLeuGlyAsnAlaProGluLysLeu 28
Db 13 CAATTATTTCCCACTTCCTTTTCCCTCAAAATTTCTTGGGATTCCTATGGCTACTGCT 72
QY 29 AlaLeuArgAsnGluGluArgAlaIleAspGluLeuLys---LysGlnAlaIleGluAsp 47
Db 73 GAGGTTGCATCAGTTCAAACTGCACCTTCATGAGGAGAAACTGAAGATCAGTGAAGGAT 132
QY 48 LysGluAlaThrThrAlaIleGluAlaIleSerSerAspAlaLeuGluAlaLeuAlaAsp 67
Db 133 CAAGAAGCTACACATGATGAGTAGTTGCCGAAGCTTCCTGCAGCGAAGAGCTGCCGAA 192
QY 68 GlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLysAlaAspAsnAlaIleAsp 87
Db 193 GAGCCAAAG-----GAAGCGAGCGCTGGCGGCTGCAGTCGAGGAACCTGAAGCCTCG 246
QY 88 AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaGluVal 107
Db 247 -----GTTGAAGTTGAAACCAAGGAGGTG 270
QY 108 ValGlnSerAspAsnAlaAlaSerAspAlaTrpGluLysAlaAlaThrProIleAlaLeu 127
Db 271 GTAGAAGAACCAAGGCTGTACTGAG-----GAGCCAGTAGTAGAG 312
QY 128 AspValLysLysThrLysAspThrLysProVal-----ValLysLysGluGluArg 144
Db 313 GAGACTCCTAAGGAACACAGTCGAGAACCAAGCTTCGCGAGGAGATCAAGGAACCACTGAA 372
QY 145 GlnAsnValAsnThrLeuProThrThr 153
Db 373 CAAACCGTTGAAACCAAGAACTACA 399

RESULT 19
BG445649
LOCUS
DEFINITION
GA_Ea0029A12f Gossypium arboreum 7-10 dpa fiber library Gossypium

```

arborescens cDNA clone GA\_Ea029A12f, mRNA sequence.

```

ACCESSION   BG445649
VERSION     BG445649.1 GI:13355301
KEYWORDS    EST.
SOURCE      Gossypium arboreum.
            Gossypium arboreum.
            Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.
REFERENCE   1 (bases 1 to 881)
AUTHORS     Wing, R.A., Frisch, D., Yu, Y., Main, D., Rambo, T., Simmons, J., Henry
            D., Wood, T.C., Leslie, A. and Wilkins, T.A.
TITLE       An integrated analysis of the genetics, development, and evolution
            of the cotton fiber
JOURNAL     Unpublished (2000)
COMMENT     Contact: Wing RA
            Clemson University Genomics Institute
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Seq primer: TAATACGACTCCTACTAGGG
            High quality sequence start: 10
            High quality sequence stop: 715.
            Location/Qualifiers
                source          1..881
                organism="Gossypium arboreum"
                strain="AKA"
                cultivar="8400"
                db_xref="taxon:29729"
                clone="GA_Ea029A12f"
                clone_lib="Gossypium arboreum 7-10 dpa fiber library"
                tissue_type="Fibers isolated from bolls harvested 7-10
                dpa"
                lab_host="E. coli"
                note="vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"

BASE COUNT   275 a 162 c 237 g 205 t 2 others
ORIGIN
Alignment Scores:
Pred. No.:    0.251      Length:      881
Score:        107.00     Matches:    39
Percent Similarity: 44.97%      Conservative: 28
Best Local Similarity: 26.17%    Mismatches: 58
Query Match:   13.77%      Indels:    24
DB:            10         Gaps:      5

US-09-847-539A-6 (1-159) x BG445649 (1-881)

Qy  9  ArgilleleProAsnGlyGlyThrLeuThrAsnLeuLeuGlyAsnAlaProGluLysLeu 28
Db  30 CAATTATTTCCCACTTCCTTTTGGCTTCCAAATTTCTTGGGATTCCTACTGCTACTGCT 89
Qy  29 AlaLeuArgAsnGluGluArgAlaLeuLeuLysLysLysGlnAlaLeuGluAsp 47
Db  90 GAGGTTGCATGAGTTCAAACTGCCTTCATGAGGAGAAACCTGAAGATCAGTGAAGAT 149
Qy  48 LysGluAlaThrAlaLeuGluAlaLeuSerSerAspAlaLeuGluAlaLeuAlaAsp 67
Db  150 CAAGAAGTACACAGATGAAGTAGTTCGCGAGGCTCTCGAGCGGAGGAGCTCCGCAA 209
Qy  68 GlnThrAspAlaLeuGlnSerGluGluAlaValValValValValValValValVal 87
Db  210 GAGCCAAAG-----GAGCGGAGCCTCGCGGGTGGCGAGCTCGGAGCACTGAAGCCTCG 263
Qy  88 AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaGluVal 107
Db  264 -----GTTGAAGTTGAACCAACCAAGGAGGTG 287
Qy  108 ValGlnSerAspAsnAlaAlaSerAspAlaTrpGluLysAlaAlaThrProIleAlaLeu 127
Db  288 GTAGAAGCAACCAAGGCTGTTACTGAG-----GAGCCAGTAGTAGAG 329

```

```

Qy  128 AspValLysLysThrLysAspThrLysProVal-----ValLysLysGluGluArg 144
Db  330 GAGACTCTTAAGAAACACAGTCCAGAACAGTTGCCGAGGAGATCAAGGAACCACTGAA 389
Qy  145 GlnAsnValAsnThrLeuProThrThr 153
Db  390 CAACCGTTGAACCAACCAAGAACTACA 416

RESULT 20
BG444821
LOCUS     BG444821
DEFINITION   BG444821
ACCESSION   BG444821
VERSION     BG444821.1 GI:13354473
KEYWORDS    EST.
SOURCE      Gossypium arboreum.
            Gossypium arboreum.
            Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.
REFERENCE   1 (bases 1 to 887)
AUTHORS     Wing, R.A., Frisch, D., Yu, Y., Main, D., Rambo, T., Simmons, J., Henry
            D., Wood, T.C., Leslie, A. and Wilkins, T.A.
TITLE       An integrated analysis of the genetics, development, and evolution
            of the cotton fiber
JOURNAL     Unpublished (2000)
COMMENT     Contact: Wing RA
            Clemson University Genomics Institute
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Seq primer: TAATACGACTCCTACTAGGG
            High quality sequence stop: 664.
            Location/Qualifiers
                source          1..887
                organism="Gossypium arboreum"
                strain="AKA"
                cultivar="8400"
                db_xref="taxon:29729"
                clone="GA_Ea025L07f"
                clone_lib="Gossypium arboreum 7-10 dpa fiber library"
                tissue_type="Fibers isolated from bolls harvested 7-10
                dpa"
                lab_host="E. coli"
                note="vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"

BASE COUNT   272 a 161 c 241 g 213 t
ORIGIN
Alignment Scores:
Pred. No.:    0.253      Length:      887
Score:        107.00     Matches:    39
Percent Similarity: 44.97%      Conservative: 28
Best Local Similarity: 26.17%    Mismatches: 58
Query Match:   13.77%      Indels:    24
DB:            10         Gaps:      5

US-09-847-539A-6 (1-159) x BG444821 (1-887)

Qy  9  ArgilleleProAsnGlyGlyThrLeuThrAsnLeuLeuGlyAsnAlaProGluLysLeu 28
Db  22 CAATTATTTCCCACTTCCTTTTGGCTTCCAAATTTCTTGGGATTCCTACTGCTACTGCT 81
Qy  29 AlaLeuArgAsnGluGluArgAlaLeuLeuLysLysLysGlnAlaLeuGluAsp 47
Db  82 GAGGTTGCATGAGTTCAAACTGCCTTCATGAGGAGAAACCTGAAGATCAGTGAAGAT 141
Qy  48 LysGluAlaThrThrAlaLeuGluAlaLeuSerSerAspAlaLeuGluAlaLeuAlaAsp 67
Db  142 CAAGAAGTACACAGATGAAGTAGTTCGCGAAGCTCTCGAGCGGAGGAGCTCCGCAA 201

```



**AUTHORS** Kohara,Y., Shin-I,T., Thierry-Mieg,J., Thierry-Mieg,D., Suzuki,Y.  
and Sugano,S.  
**TITLE** A complementary view of the C.elegans genome  
**JOURNAL** Unpublished (2002)  
**COMMENT** Contact: Tadasi Shin-I  
Center For Genetic Resource Information  
National Institute of Genetics  
1111 Yata, Mishima, Shizuoka 411-8540, Japan  
Tel: 81-559-81-6856  
Fax: 81-559-81-6855  
Email: tshin@genes.nig.ac.jp.

**FEATURES** Location/Qualifiers  
source  
1. .620  
/organism="Caenorhabditis elegans"  
/strain="N2"  
/db\_xref="taxon:6239"  
/clone\_lib="y1183b11"  
/clone\_lib="unpublished oligo-capped cDNA library, C.  
elegans L1 stage"  
/sex="hermaphrodite"  
/tissue\_type="whole animal"  
/dev\_stage="L1"  
200 a 150 c 141 g 129 t

**BASE COUNT** 200 a 150 c 141 g 129 t  
**ORIGIN**

**Alignment Scores:**

Pred. No.:	0.21	Length:	620
Score:	106.00	Matches:	41
Percent Similarity:	45.03%	Conservative:	27
Best Local Similarity:	27.15%	Mismatches:	59
Query Match:	13.64%	Indels:	24
DB:	10	Gaps:	6

US-09-847-539a-6 (1-159) x BJ114733 (1-620)

Qy 17 LeuThrAsnLeuLeuGlyAsnAlaProGluLysLeu-----AlaLeuArg 31  
||| ||| :|||:||||| :||| :|||  
Db 68 CTGAAGAACGATGTCGGAATGCGGTCAAGACGCGGAAGATGTTGTTTCGNAATACCAA 127  
Qy 32 AsnGluGluArgAlaIleAspGluLeuLysGlnAlaIleGluAspLysGluAlaThr 51  
||||| :||| :|||:||||| :||| :|||  
Db 128 AATCAACCTCAACCACTCGACGCTGGCTAAGGACGCGCTAACAACTCAAG----- 178

Qy 52 ThrAlaIleGluAlaAlaSerSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAla 71  
||| :||| :||| :|||:||||| :||| :|||  
Db 179 -----GCTACTGTAGAGCACTTACCAGCTGGCGGCTAGCGACAAG 223

Qy 72 LeuGlnSerGluGluAlaAlaValValLysAlaAspAsnAlaAlaSer----- 87  
||| :||| :||| :|||:||||| :||| :|||  
Db 224 ATCGATCCACAAGTCGCCAAGGACATCAAGGACTCAAGAGCTAGGCTAAGGAATTGCTT 283

Qy 88 AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaGluVal 107  
||||| ||| ||| |||:||||| :||| :|||  
Db 284 CAGGCTCTTGAAAGCCATTCCACAGAGGACGCTATTCGTCGTGAACAAGCCGAGATT 343

Qy 108 ValGlnSerAspAsnAlaAlaSerAspAlaTrpGluLysAlaAlaThrProfileAlaLeu 127  
||| :||| :||| :|||:||||| :||| :|||  
Db 344 -----AATGATCGTCTCAACAATCTCGAAGAGGAGCTTACCAGGTG----- 385

Qy 128 AspValLysLysThrLysAspThrLysProValValLysLysGluGluArgGlnAsnVal 147  
||| ||| :||| :|||:||||| :||| :|||  
Db 386 GATGAGTTCAACACGAGGATGCATCTTCCATTTGTT---GACCAACTCGCCGCCAATACT 442

Qy 148 AsnThrLeuProThrThrGlyGluGluSerAsn 158  
||||| ||| :||| :|||:||||| :||| :|||  
Db 443 AACACCTTGAAGACTGCCACTGATTTCGAACAAC 475

RESULT 23  
LOCUS CNS0794Y 947 bp DNA linear GSS 08-JUL-2001  
DEFINITION T3 end of clone BB0AA026C12 of library BB0AA from strain CBS 4732  
of Pichia angusta, genomic survey sequence.  
ACCESSION AL434840

AL434840.1 GI:12218253  
GSS.  
Pichia angusta.  
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
Saccharomycetales; Saccharomycetaceae; Pichia.  
1 (bases 1 to 947)  
Blandin,G., Llorente,B., Malpertuy,A., Wincker,P., Artiguenave,F.  
and Dujon,B.  
Genomic exploration of the hemiascomycetous yeasts: 13. Pichia  
angusta  
FEBS Lett. 487 (1), 76-81 (2000)  
20584723  
2 (bases 1 to 947)  
Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,  
Boletín-Pukuhara,M., Bon,E., Brottier,P., Casaregola,S.,  
de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,  
Malpertuy,A., Neuvéglise,C., Ozler-Kalogeropoulos,O., Potier,S.,  
Saurin,W., Tekala,F., Toffano-Nloche,C., Wesolowski-Louvel,M.,  
Wincker,P. and Weissbach,J.  
Genomic exploration of the hemiascomycetous yeasts: 1. A set of  
yeast species for molecular evolution studies  
FEBS Lett. 487 (1), 3-12 (2000)  
20584711  
3 (bases 1 to 947)  
Genoscope.  
Direct Submission  
Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,  
2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :  
seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)  
This GSS is part of a random genomic sequencing program of thirteen  
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces  
exiguus, Saccharomyces servazii, Zygosaccharomyces rouxii,  
Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces  
lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia  
angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,  
Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to  
5 kb were prepared and both extremities were sequenced. See  
keywords for description of this sequence and for the sequence of  
the other extremity of this insert.

**FEATURES** Location/Qualifiers  
source  
1. .947  
/organism="Pichia angusta"  
/strain="CBS 4732"  
/db\_xref="taxon:4905"  
/clone\_lib="BB0AA026C12"  
/clone\_lib="BB0AA"  
/note="end : T3"  
280 a 198 c 293 g 175 t 1 others

**BASE COUNT** 280 a 198 c 293 g 175 t 1 others  
**ORIGIN**

**Alignment Scores:**

Pred. No.:	0.346	Length:	947
Score:	106.00	Matches:	41
Percent Similarity:	50.86%	Conservative:	18
Best Local Similarity:	35.34%	Mismatches:	44
Query Match:	13.64%	Indels:	13
DB:	12	Gaps:	5

US-09-847-539a-6 (1-159) x CNS0794Y (1-947)

Qy 7 GlnProArgIleIleProAsn-----GlyGlyThrLeuThrAsnLeuLeuGlyAsn-Al 24  
||| ||| :||| :||| :|||:||||| :||| :|||  
Db 433 CAGACGAGAGCTCTGGGAACCTCTCTGGGAGGAGCTCAGTAGCAGCAGCAATAGC 492

Qy 24 aProGluLysLeuAlaLeuArgAsnGlu-----GluArgAlaIleAs 38  
||| ||| :||| :|||:||||| :||| :|||  
Db 493 TAGCAGACCTTGGCCTTCTCCACGAAGAGAGACTCAGCAGCAAGGAGGCTGCAGA 552

Qy 38 pGluLeuLysGlnAlaIleGluAspLysGluAlaThrAlaIleGluAlaIleSe 58  
||| :||| :||| :|||:||||| :||| :|||  
Db 553 CGCGCGGAAGCAGAGGCTTCGGAGCGCTCAGAA---ACAGAGCAGCAGCAAGCTGCCGC 609



```

Db 160 GATGCTGGGACAAAAGAGCCAGCTGTTACAGAACGAGGTGAACAGCTGAACCTTCAGCT 219
Qy 22 GlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaIleAspGluLeuLys 41
Db 220 GATTTCATCTGACAGTGAATGCTATGATGACAACTCT-----GAGGACAACAGC 270
Qy 42 LysGlnAlaIleGluAspLysGluAlaThrAlaIleGluAlaAlaSerSerAspAla 61
Db 271 GATGCATCTATGATGGACAATCTGAAGACAACGCTCAAGACGCTCCACTGAACATGATGAA 330
Qy 62 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLys 81
Db 331 GAAAAACCAAGAAATGACCAACTGACACACAGAGGAGGATTCACGATGAAGAAC 390
Qy 82 AlaAspAsnAlaAlaSerAspAlaLeuGluAlaLeuAla----- 94
Db 391 TCTGACGATGAGTCCAAAGAAACATCTAAAGCTGCTGCAGAGTCAAGACCACTGAGAAA 450
Qy 95 -----AspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLys 109
Db 451 ACAGCTGAAACCCAGACGCTACAGATGAACCAAGTGAAGAAACATCAGAGGCGAGGAA 510
Qy 110 SerAspAsnAlaAlaSerAspAlaTTPGluLysAlaAlaThrProIleAlaLeuAspVal 129
Db 511 TTGGAT---TCAGCATGACGACCAAGCTCTACAGATGAGCCA---AGTCAAGAAACA 564
Qy 130 LysLysThrLysAspThrLysProValValLysLysGluGluArgGlnAsnValAsnThr 149
Db 565 GCAGAGACTAAAGAACAGAGTCAGCAGATAAAGTGAAGCAACAAAGT-----TCAGCA 618
Qy 150 LeuProThrThrGlyGluGluSerAsn 158
Db 619 GCAGAGAAACTGATGAAGAAGCCTCA 645

RESULT 26
BJ011593 687 bp mRNA linear EST 05-DEC-2001
LOCUS BJ011593
DEFINITION BJ011593 MF01SSA cDNA Oryzias latipes cDNA clone MF01SSA165C04 5',
mRNA sequence.
ACCESSION BJ011593
VERSION BJ011593
KEYWORDS EST.
SOURCE Japanese medaka.
ORGANISM Oryzias latipes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percormorpha; Atherinomorpha;
Belontiiformes; Adrianichthyidae; Oryziinae; Oryzias.
1 (bases 1 to 687)
Kohara,Y., Shin-I.,T., Kimura,T., Narita,T., Jindo,T. and Takeda,H.
Medaka EST Project in Takeda's lab
Unpublished (2001)
Contact: Tadasu Shin-I
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.

FEATURES
    source
    1..687
    /organism="Oryzias latipes"
    /strain="Hd-rR"
    /db_xref="taxon:8090"
    /clone="MF01SSA165C04"
    /clone_lib="MF01SSA cDNA"
    /sex="mixture of female and male"
    /tissue_type="whole embryo"
    /dev_stage="segmentation stage 20 - 25"

BASE COUNT 233 a 143 c 180 g 131 t
ORIGIN
Alignment Scores:

```

```

Pred. No.: 0.267 Length: 687
Score: 105.50 Matches: 34
Percent Similarity: 44.97% Conservative: 42
Best Local Similarity: 20.12% Mismatches: 74
Query Match: 13.58% Indels: 19
DB: 5

US-09-847-539A-6 (1-159) x BJ011593 (1-687)
Qy 2 AspSerProIleGluGlnProArgIleIleProAsnGlyGlyThrLeuThrAsnLeuLeu 21
Db 170 GATGCTGGGACAAAAGAGCCAGCTGTTACAGAACGAGGTGAACAGCTGAACCTTCAGCT 229
Qy 22 GlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaIleAspGluLeuLys 41
Db 230 GATTTCATCTGACAGTGAATGCTATGATGACAACTCT-----GAGGACAACAGC 280
Qy 42 LysGlnAlaIleGluAspLysGluAlaThrAlaIleGluAlaAlaSerSerAspAla 61
Db 281 GATGCATCTATGATGGACAATCTGAAGACAACGCTCAAGACGCTCCACTGAACATGATGA 340
Qy 62 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLys 81
Db 341 GAAAAACCAAGAAATGACCAACTGACACACAGAGGAGGATTCACGATGAAGAAC 400
Qy 82 AlaAspAsnAlaAlaSerAspAlaLeuGluAlaLeuAla----- 94
Db 401 TCTGACGATGAGTCCAAAGAAACATCTAAAGCTGCTGCAGAGTCAAGACCACTGAGAAA 460
Qy 95 -----AspGlnThrAspAlaLeuGlnSerGluGluAlaGluValValGln 109
Db 461 ACAGCTGAAACCCAGACGCTACAGATGAACCAAGTGAAGAAACATCAGAGGCGAGGAA 520
Qy 110 SerAspAsnAlaAlaSerAspAlaTTPGluLysAlaAlaThrProIleAlaLeuAspVal 129
Db 521 TTGGAT---TCAGCATGACGACCAAGCTCTACAGATGAGCCA---AGTCAAGAAACA 574
Qy 130 LysLysThrLysAspThrLysProValValLysLysGluGluArgGlnAsnValAsnThr 149
Db 575 GCAGAGACTAAAGAACAGAGTCAGCAGATAAAGTGAAGCAACAAAGT-----TCAGCA 628
Qy 150 LeuProThrThrGlyGluGluSerAsn 158
Db 629 GCAGAGAAACTGATGAAGAAGCCTCA 655

RESULT 27
BJ001283
LOCUS BJ001283
DEFINITION BJ001283 MF01SSA cDNA Oryzias latipes cDNA clone MF01SSA005B12 5',
mRNA sequence.
ACCESSION BJ001283
VERSION BJ001283
KEYWORDS EST.
SOURCE Japanese medaka.
ORGANISM Oryzias latipes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percormorpha; Atherinomorpha;
Belontiiformes; Adrianichthyidae; Oryziinae; Oryzias.
1 (bases 1 to 710)
Kohara,Y., Shin-I.,T., Kimura,T., Narita,T., Jindo,T. and Takeda,H.
Medaka EST Project in Takeda's lab
Unpublished (2001)
Contact: Tadasu Shin-I
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.

FEATURES
    source
    1..710
    /organism="Oryzias latipes"

```







```
Query Match: 13.51% Indels: 12
DB: 9 Gaps: 4

US-09-847-539A-6 (1-159) x AJ395963 (1-755)

QY 17 LeuThrAsnLeuLeuGlyAsnAlaProGluLysLeuAla-----LeuArgAsnGluGlu 34
   ||| ||| ::| ||||| ::| ::| ::| ::| ::| ::| ::| ::| ::|
Db 85 TTGGAGAATAAATGAAATGCAAGACAGCAGTGGGAAGGAAATAAATAATGCCAG 144
   ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
QY 35 AtqAlaileAspGluLysLysGlnAlaIleGluAspLysGluAlaThrAlaIle 54
   ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
Db 145 CAGAACTAAATAGCCCAAGAGAAAGCAGGATTCAGCAGAAAAATG----- 195
   ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
QY 55 GluAlaAlaSerSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSer 74
   ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
Db 196 ---AAAGAAAGCAGCAAGAGTTGAAGCATTAGTCTGGAAGTTCAGACAGCTGAACAA 252
   ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
QY 75 GluGluAlaAlaValLysAlaAspAsnAlaAlaSerAspAlaLeuGluAlaLeuAla 94
   ||| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
Db 253 GAACAAGCCTCATATAGCAACAGAGTGAGCGACACAAACAGCAATCGCATCTTAAAG 312
   ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
QY 95 AspGlnThrAspAlaLeuGlnSerGluGluAlaGluValValGlnSerAspAsnAlaAla 114
   ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
Db 313 GACCAAGTCAGTCTTG-----GAAGCTGAGCGAGTGAAGACAGGGAATCTCTA 363
   ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
QY 115 SerAspAlaTrpGluLysAlaAlaThrProIleAlaLeuAspValLysThrLysAsp 134
   ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
Db 364 AAGAAGCAGAGATGAGCTGCTCCATGAAAGGGATTATGCGCAGACGAACTAAGAT 423
   ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
QY 135 ThrLys-----ProValValLysLysGluGluArgGlnAsn 146
   ||| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
Db 424 ATTAAGCCAAATCTGCAAGATAGAGAAATACAGAGAGCAAAAT 468

RESULT 31
BM300555
LOCUS BM300555 530 bp mRNA linear EST 22-JAN-2002
DEFINITION MCA053H02.21480 Ice plant Lambda Uni-zap XR expression library, 0
           hours NaCl treatment prescreened for removal of highly abundant
           transcripts Mesembryanthemum crystallinum cDNA clone MCA053H02 5,
           mRNA sequence.
ACCESSION BM300555
VERSION BM300555.1 GI:18022930
KEYWORDS EST.
SOURCE common ice plant.
ORGANISM Mesembryanthemum crystallinum
           Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
           Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
           Caryophyllidae; Caryophyllales; Aizoaceae; Mesembryanthemum.
           1 (bases 1 to 530)
           Cushman, J.C.
REFERENCE An expressed sequence tag database for the common ice plant,
           Mesembryanthemum crystallinum
           Unpublished (1997)
JOURNAL Contact: Cushman JC
COMMENT Department of Biochemistry
           University of Nevada
           MS200, Reno, NV 89557-0014, USA
           Tel: 775-784-1918
           Fax: 775-784-1650
           Email: jcushman@unr.edu
PCR PRIMERS
FORWARD: T3 20mer
BACKWARD: T7 21mer
Plate: 053 row: H column: 02
Seq primer: T3 20mer
High quality sequence stop: 530.
Location/Qualifiers
1..530
/organism="Mesembryanthemum crystallinum"
/db_xref="taxon:3544"
/clone="MCA053H02"
, 0 hours NaCl treatment prescreened for removal of highly

abundant transcripts"
/tissue_type="leaf"
/dev_stage="six-week-old"
/note="Vector: Lambda Uni-zap XR, Bluescript SK-; Site_1:
EcoRI; Site_2: XhoI"
BASE COUNT 186 a 99 c 173 g 72 t
ORIGIN
Alignment Scores:
Pred. No.: 0.28 Length: 530
Score: 104.00 Matches: 44
Percent Similarity: 40.13% Conservatives: 17
Best Local Similarity: 28.95% Mismatches: 49
Query Match: 13.38% Indels: 42
DB: 10 Gaps: 7

US-09-847-539A-6 (1-159) x BM300555 (1-530)
QY 30 LeuArgAsnGluArgAlaIleAspGluLysLysGlnAlaIleGluAspLysGlu 49
   ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 9 TTGACGAACGAGGAGGGGCG-----GCCGCGCGCGCGGAGGAAGCCA 53
   ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
QY 50 AlaThrAlaIleGluAlaAlaSerSerAspAlaLeuGluAlaLeuAlaAspGlnThr 69
   ||| ||| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
Db 54 GCA-----GCGGTGAAGTTGCGCGCGGGAAGTGGCGAAGTGTGGCGGAAAGTAGA 107
   ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
QY 70 AspAla-----LeuGlnSerGluGluAlaAlaValValLysAlaAspAla 85
   ||| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
Db 108 GATATTCGGTGTGACTTTAAAAAGAGGAGGAAGAAAGAACGCCGAGCTGAGGAGGCC 167
   ||| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
QY 86 AlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAla 105
   ||| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
Db 168 GCAAG-----CCAACCGAGGAGAG 188
   ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
QY 106 GluValValGlnSerAspAsnAlaAlaSerAspAlaTrpGluLysAlaAlaThrProIle 125
   ::| ::| ||||| ||| ||| ||||| ||||| |||
Db 189 GCTGCAGAGCCAGCCAAAGAAAGCCGAAAGCCAGCAGAGGAAAGGCTGTGAGCCAGCC 248
   ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
QY 126 AlaLeuAspValLysThrLysAsp-----Thr 135
   ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
Db 249 ACCAAGAAGCTGAGACCAAGAAAGAAAGAGTCCCCAAGCCAGCTGAAGCTGAACCGGCT 308
   ||| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
QY 136 LysProValVal-----LysLysGluGluArgGlnAsnValAsnThrLeuProThr 152
   ||||| ||| ||||| ||||| ::| |||
Db 309 AAGCCAGTACCCCGCCGAGCAAGAGAGGAGGAGCAAGCCAGCTTGAGGAGCTCCAGCA 368
   ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
QY 153 ThrGly-----GluGluSerAsnPro 159
   ||||| ::| ::|
Db 369 GCGGCTGCTGAGAAGAAGGTTGAGGAGAGCAAAACCC 404
   ||||| ::| ::|

RESULT 32
BJ006142
LOCUS BJ006142 616 bp mRNA linear EST 05-DEC-2001
DEFINITION BJ006142 MF01SSA cDNA Oryzias latipes cDNA clone MF01SSA089B12 5',
           mRNA sequence.
ACCESSION BJ006142
VERSION BJ006142.1 GI:17359859
KEYWORDS EST.
SOURCE Japanese medaka.
ORGANISM Oryzias latipes
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
           Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
           Belontiiformes; Adrianichthyidae; Oryziinae; Oryzias.
           1 (bases 1 to 616)
           Kohara, Y., Shin-i, T., Kimura, T., Narita, T., Jindo, T. and Takeda, H.
           Medaka EST Project in Takeda's lab
           Unpublished (2001)
           Contact: Tadasu Shin-i
           Center For Genetic Resource Information
           National Institute of Genetics
           1111 Yata, Mishima, Shizuoka 411-8540, Japan
           Tel: 81-559-81-6856
REFERENCE
AUTHORS
JOURNAL
COMMENT
```



ORGANISM Entamoeba histolytica  
Eukaryota; Entamoebidae; Entamoeba.

REFERENCE 1 (bases 1 to 938)

AUTHORS Loftus B., Van Aken S. and Fraser C.

TITLE Determination of clone end sequences from Entamoeba histolytica

JOURNAL HMI:IMSS sheared DNA library

COMMENT Unpublished (2000)  
Contact: Brendan J Loftus  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0208  
Fax: 301 838 3543  
Email: bjlouftus@tigr.org  
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared DNA library  
Seq primer: M13-Reverse  
Class: shotgun  
High quality sequence start: 34  
High quality sequence stop: 809.

FEATURES  
source  
1..938  
/organism="Entamoeba histolytica"  
/strain="HMI:IMSS"  
/db\_xref="taxon:5759"  
/clone\_lib="Entamoeba histolytica Sheared DNA"  
/note="Vector: pHS1; Site\_1: Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Borell, Oxford University Press, 1999)."

BASE COUNT 132 a 201 c 107 g 498 t

ORIGIN  
Alignment Scores:  
Pred. No.: 0.697 Length: 938  
Score: 103.00 Matches: 40  
Percent Similarity: 49.62% Conservative: 26  
Best Local Similarity: 30.08% Mismatches: 59  
Query Match: 13.26% Indels: 8  
DB: 12 Gaps: 4

US-09-847-539A-6 (1-159) x AZ676392 (1-938)

Qy 16 ThrLeuThrAsnLeuLeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArg 35  
:::|||||  
Db 829 AGTCTAAGAAGAGCTGAAGCAAAAGCCAAAGAAAGAGCT-----GAAGCAAAA 779

Qy 36 AlaIleAspGluLeuLysLysGlnAlaIleGluAspLysGluAlaThrAlaIleGlu 55  
:::|||||  
Db 778 GCTTAAGAAGAGCTGAAGCAAAAGCTTAAAGAAGAGAGCTGAAGCAAAAGCTAAAGAAGAA 719

Qy 56 AlaAlaSerSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGlu 75  
|||  
Db 718 GCTGAACCAAAAGCTTAAAGAAGAGCTGAAGCAAAAGCCAAAGAA-----GAAGCTGAA 665

Qy 76 GluAlaAlaValValLysAlaAspAsnAlaAlaSerAspAlaLeuGluAlaLeuAla--- 94  
|||  
Db 664 GCAAAAGCTTAAAGAAGAGCTGAAGCAAAAGCCAAAGAAAGCTGAAGCAAAAGCTTAA 605

Qy 95 AspGlnThrAspAlaLeuGlnSerGluGluAlaValValGlnSerAspAsnAlaAla 114  
:::|||||  
Db 604 GAAGAGCTGAAGCAAAAGCTTAAAGAAGAGCTGAAGCAAAAGCTTAAAGAAGAGCTGAA 545

Qy 115 SerAspAlaTrpGluLysAlaAlaThrProIleAlaLeuAspValLysLysThrLysAsp 134

Db 544 GCAAAAGCTAAGAAGAGCTGAAGCAAAAGCTTAAAGAAGAGCTTAA-----GAAGAA 491  
::: ||| |||:::|||||  
Qy 135 ThrLysProValValLysLysGluGluArgGlnAsnVal 147  
||| |||:::|||||  
Db 490 AATAAAGAAGAGCTTAAAGAAGAGCTTAAAGAAGAGCTT 452  
RESULT 35  
BG930118 537 bp mRNA linear EST 31-DEC-2001  
LOCUS etSHEST0539 Eth11 Eimeria tenella cDNA clone etsHk077 5', mRNA  
DEFINITION sequence.  
ACCESSION BG930118  
VERSION BG930118 GI:18003508  
KEYWORDS EST.  
SOURCE Eimeria tenella.  
ORGANISM Eimeria tenella.  
REFERENCE 1 (bases 1 to 537)  
AUTHORS Ng, S.T., Jangi, M.S., Shirley, M.W., Tomley, F.M. and Wan, K.L.  
TITLE Comparative EST analyses provide insights into gene expression in two asexual developmental stages of Eimeria tenella  
JOURNAL Unpublished (2001)  
COMMENT Contact: Wan KL  
Centre for Gene Analysis and Technology  
Universiti Kebangsaan Malaysia  
43600 UKM Bangi, Selangor DE, Malaysia  
Tel: 6 03 8292997  
Fax: 6 03 8293249  
Email: kluan@krisc.cc.ukm.my  
PCR Primers  
FORWARD: T3  
BACKWARD: T7  
Seq primer: SK.

FEATURES  
source  
1..537  
/organism="Eimeria tenella"  
/strain="Houghton"  
/db\_xref="taxon:5802"  
/clone\_lib="etsHk077"  
/dev\_stage="Sporozoite"  
/lab\_host="XLI-Blue MRF"  
/note="Vector: Lambda ZAPII; Site\_1: EcoRI; Site\_2: XhoI; Sporozoites were excysted in vitro from E. tenella H oocysts, purified by column chromatography and mRNA extracted using a FASTTRACK kit (Invitrogen). cDNA was synthesised and a Uni-ZAP XR library was constructed using cDNA synthesis kit, ZAP-cDNA synthesis kit and ZAP-cDNA Gigapack III Gold cloning kit (Stratagene). The library was amplified once through E. coli XLI-Blue MRF'."

BASE COUNT 175 a 140 c 191 g 29 t 2 others

ORIGIN  
Alignment Scores:  
Pred. No.: 0.405 Length: 537  
Score: 102.50 Matches: 42  
Percent Similarity: 42.42% Conservative: 28  
Best Local Similarity: 25.45% Mismatches: 61  
Query Match: 13.19% Indels: 34  
DB: 10 Gaps: 5

US-09-847-539A-6 (1-159) x BG930118 (1-537)

Qy 29 AlaLeuArgAsnGluGluArgAlaIleAspGluLeuLysLysGlnAlaIleGluAspLys 48  
:::|||||  
Db 8 GCACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 67

Qy 49 GluAlaThrThrAla-----IleGluAlaAlaSerSerAspAlaLeuGlu----- 63  
|||||  
Db 68 GAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 127

QY 64 -----AlaLeuAlaAspGln 68  
 Db 128 CGTCACGCTCGGGTGAAGACGAGCTGAAGAGAGAAAGCGGCACGACTCT 187  
 QY 69 ThrAsp-----AlaLeuGlnSerGluAlaAlaValValLysAlaAspAsn 84  
 Db 188 GGAGAGAGAGACGACGACGAGGAGCGGAGAGAGAGAGAGAGAGAGAGAG 247  
 QY 85 AlaAlaSerAspAlaLeuAlaLeuAlaAspGln-----ThrAspAlaLeu 100  
 Db 248 CGCGACGAGAGTTCGGAAGACGACGAGCTCAGGGCGCGCGACGACGACGA 307  
 QY 101 GlnSerGluAlaAlaGluValGlnSerAspAsn-----AlaAlaSer 115  
 Db 308 GAGTCGAGAGAGAGACGACGACGAGAGACTCTGACGACGACGACGACGAG 367  
 QY 116 AspAlaTrpGluLysAlaAlaThrProLeu-AlaLeuAspValLysThrLysAspTh 135  
 Db 368 GACTCTGAGGAGAGACGACGACGAGAGTCCCGGATGCGACGACGACGAGC 427  
 QY 135 rLysProValValLysLysGluGluArgGlnAsnValAsnThrLeuProThrThrGlyGl 155  
 Db 428 GAGGACCGGACGACGACGACGACGAGTCCCGGATGCGACGACGACGAGTCA 487  
 QY 155 uGluSerAsnPro 159  
 Db 488 GCAGCAGACTCCA 500

RESULT 36  
 A0648620  
 LOCUS  
 DEFINITION  
 RPCI93-EcoRI-6D15, TV RPCI93-EcoRI Trypanosoma brucei genomic clone  
 RPCI93-EcoRI-6D15, DNA sequence.

ACCESSION  
 A0648620  
 VERSION  
 GSS.  
 KEYWORDS  
 Trypanosoma brucei.  
 SOURCE  
 Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;  
 Trypanosoma.  
 1 (bases 1 to 628)  
 EL-Sayed, N., Zhao, S., Zhao, H., Gill, S., Suh, E., Malek, J., Fujil, C.,  
 Gerard, C., Leech, V., de Jong, P., Ullu, E., Melville, S., Donelson, J.,  
 Fraser, C. and Adams, M.  
 Use of BAC end sequences from Trypanosoma brucei GUTat 10.1 RPCI-93  
 Library for gene discovery and sequence-ready map construction  
 Unpublished (1999)  
 Other\_GSS: RPCI93-EcoRI-6D15.TJ  
 Contact: Najib M. El-Sayed  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: nelsayed@tigr.org

Clones and high density filters may be purchased from BACPAC  
 Resources (<http://bacpac.med.buffalo.edu>). BAC end sequences search  
 page: <http://www.tigr.org/ldb/ldb/tbdb/>.  
 Seq primer: 17  
 Class: BAC ends.

FEATURES  
source

Location/Qualifiers  
 1..628  
 /organism="Trypanosoma brucei"  
 /strain="TREU927/4 GUTat 10.1"  
 /db\_xref="taxon:5691"  
 /clone\_lib="RPCI93-EcoRI"  
 /note="Vector: pBACE3.6; Site.1: Eco RI; Site.2: Eco RI;  
 Constructed for The Institute for Genomic Research by  
 Bohui Zhao in Pieter de Jong's laboratory (Roswell Park  
 Cancer Institute, Buffalo, NY). Briefly, Trypanosoma  
 brucei TREU927/4 GUTat 10.1 agarose embedded DNA was

partially digested with a combination of Eco RI and Eco RI  
 methylase (RPCI93-EcoRI segment) or Dpn II (RPCI93-DpnII  
 segment). High molecular weight fragments were ligated in  
 pBACE3.6 vector digested with Eco RI or Bam HI,  
 respectively. The average insert size is 141 Kb. Total  
 coverage (both segments): > 90 X the haploid  
 non-minichromosomal genome."

BASE COUNT 218 a 127 c 168 g 114 t 1 others  
 ORIGIN

Alignment Scores:  
 Pred. No.: 0.488 Length: 628  
 Score: 102.50 Matches: 47  
 Percent Similarity: 44.44% Conservative: 33  
 Best Local Similarity: 26.11% Mismatches: 61  
 Query Match: 13.19% Indels: 39  
 DB: 12 Gaps: 8

US-09-847-539A-6 (1-159) x A0648620 (1-628)

QY 1 ValAspSerProGluGlnProArgGlnProAsnGlyGlyThrLeuThrAsnLeu 20  
 Db 39 GTCTGAAGATAGCTCGCTGAACCCCGT-----CGTCTTCTGAAGAA 80  
 QY 21 LeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArg----- 35  
 Db 81 ATAGACAACCGTGTGCAAAATTTGAAGGTGACTGTGGATGAAAAAATGCGGAGAAGTCA 140  
 QY 36 ---AlaLeuAspGluLysGlnAlaLeuGluAspLysGluAlaThrThrAlaLeu 54  
 Db 141 GCTGCTGAGGATGAGCTTCACAAACGGAAGAAGAACTCGCTCAACGACACGAGGAGCGT 200  
 QY 55 GluAlaAlaSerSer-----AspAlaLeuGluAlaLeu 65  
 Db 201 GAGGACCGCCGTACAAAAGACAGCGTTCAGAGAGGAGTGGATCGTTGAAATCCGGG 260  
 QY 66 AlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValLysAlaAspAsnAla 85  
 Db 261 GCCGATGACGACGTCGAGTCAAGGAAGAAAGAAAGAAACAAATAA-----GATGCT 314  
 QY 86 AlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSer----- 102  
 Db 315 GCAAGCGACATCAAAAGGCTCAACTACAAACAGGAGCGGATCGAGAGGCTGCCATC 374  
 QY 103 -----GluGluAlaGluValGlnSerAsp---Asn 112  
 Db 375 CATCAGCAGATCTCAATGAAGCGCGAGAACAAAGTAGAAAAACTTCAGAAAGAGTATGAC 434  
 QY 113 AlaAlaSerAspAlaTrpGluLysAlaAlaThrProLeuAlaLeuAspVal---LysLys 131  
 Db 435 GTCCGCAACAGAGAGGTTTGTAGCATCTCACACCACCTTCGCCAGGAGCTAGATAGAAG 494  
 QY 132 ThrLysAspThrLysProValValLysLys-----GluGluArgGlnAsnValAsnThr 149  
 Db 495 AAGGCTGATTTTCCCCCATATGAAAAAGCTCTGTTGGAAGCAAGAGGACCACTTGATACG 554

RESULT 37  
 CNS07270 790 bp DNA linear GSS 07-JUL-2001  
 LOCUS  
 clone BAOAB008E01 of library BAOAB from strain CLIB 210 of  
 Kluyveromyces lactis, genomic survey sequence.

ACCESSION  
 AL425866  
 VERSION  
 AL425866.1 GI:12209060

KEYWORDS  
 GSS.  
 SOURCE  
 Kluyveromyces lactis.

ORGANISM

REFERENCE  
 1 (bases 1 to 790)  
 Saccharomycetales; Ascomycota; Saccharomycetaceae; Kluyveromyces.

AUTHORS  
 Bolotin-Fukuhara, M., Toffano-Nioche, C., Artiguenave, F.,  
 Duchateau-Nguyen, G., Lemaire, M., Marmelisse, R., Montrocher, R.,  
 Robert, C., Termier, M., Wincker, P. and Wesolowski-Louvel, M.  
 Genomic exploration of the hemiascomycetous yeasts: 11.



```

Db 281 GTAGAAGAACCAAGGCTGTTACTGAGGAGCCAGTAGTAGAGGAGACTCCTTAAGGAACA 340
Qy 121 AlaAlaThrProIleAlaLeuAspVallys-----LysThrLys 133
Db 341 GTGCCAGAACCAGTTGCCGAGGAGATCAAGGAACACCTGAACANACCCGTTGAACACCAA 400
Qy 134 AspThr-----LysProValVallysGlu 142
Db 401 GAAACTACAGACTGTAGTGAGGCAGCAGCAGAGCACCACCAAGAGAGAGTTGAGAAGAA 460
Qy 143 Glu 143
Db 461 GAA 463

RESULT 39
LOCUS      CNS06MOG      1076 bp      DNA      linear      GSS 17-JUN-2001
DEFINITION T3 end of clone AU0AA006A09 of library AU0AA from strain CBS 3082
of Saccharomyces kluyveri, genomic survey sequence.
ACCESSION  AL405734
VERSION     AL405734.1  GI:12168821
KEYWORDS   GSS.
SOURCE     Saccharomyces kluyveri.
ORGANISM   Saccharomyces kluyveri.
REFERENCE  Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
AUTHORS    Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
1 (bases 1 to 1076)
Neueglise,C., Bon,E., Lepingle,A., Wincker,P., Artiguenave,F.,
Gaillardin,C. and Casaregola,S.
Genomic exploration of the hemiascomycetous yeasts: 9.
Saccharomyces kluyveri
FEMS Lett. 487 (1), 56-60 (2000)
20584719
2 (bases 1 to 1076)
Souclet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
Botolin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Liorente,B.,
Malpertuy,A., Neueglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
Saurin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
Wincker,P. and Weissenbach,J.
Genomic exploration of the hemiascomycetous yeasts: 1. A set of
yeast species for molecular evolution studies
FEMS Lett. 487 (1), 3-12 (2000)
20584711
3 (bases 1 to 1076)
Genoscope.
Direct Submission
Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,
2 rue Gaston Cremieux, Cp 5706, 91057 EVRY cedex, FRANCE. (E-mail :
seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
This GSS is part of a random genomic sequencing program of thirteen
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
5 kb were prepared and both extremities were sequenced. See
keywords for description of this sequence and for the sequence of
the other extremity of this insert.
FEATURES             source
    location/Qualifiers
        1..1076
            /organism="Saccharomyces kluyveri"
            /strain="CBS 3082"
            /db_xref="taxon:4934"
            /clone="AU0AA006A09"
            /clone_lib="AU0AA"
            /note="end : T3"

BASE COUNT      221 a 290 c 180 g 383 t      2 others
ORIGIN

Alignment Scores:      1.04      Length:      1076
Pred. No.:

```

```

Score:      102.00      Matches:      35
Percent Similarity: 50.81%      Conservative: 28
Best Local Similarity: 28.23%      Mismatches: 49
Query Match: 13.13%      Indels: 12
DB: 12      Gaps: 4

US-09-847-539a-6 (1-159) x CNS06MOG (1-1076)
Qy 23 AsnAlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaIleAspGluLeuLysLys 42
Db 832 GATGCTGAAGAAACCCGAAAGAGCGGAGACCAAGAGAGCTTCATTGGCAGCAGATTCTAGG 773
Qy 43 GlnAlaIle---GluAspLysGluAlaThrAlaIleGluAlaAala-----Ser 58
Db 772 GAAGCACTAGCGGAGAGCAGAGAGAGCTGAGCTACCGAAGCGGCACCTGGAAGAGAGAT 713
Qy 59 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAla 78
Db 712 TGTGAGGCTGCTACCGCGCTGACAGAAGATGCGGACATTTCTAGAAGCTACTGAAGCTTCA 653
Qy 79 ValVal-----LysAlaAspAsnAlaAlaSerAspAlaLeuGluAlaLeuAla 94
Db 652 CTGTTAGCAGAACCCAGCAGACGAGCAGCAGAGCAGAGCAGCAGCAGAGCA 593
Qy 95 AspGlnThrAspAlaLeuGlnSerGluGluAla-----GluValValGlnSer 110
Db 592 GCGGAAGCAGAAACCAACAGCAACAGCAACAGCAACAGCAACAGCAACAGCG 533
Qy 111 AspAsnAlaAlaSerAspAlaThrPGLuLysAlaAlaThrProIleAlaLeuAspVallys 130
Db 532 GATGACGCGCTAGCAGAGCAGCGGAATCAGAAAGCAACAGCAACAGCAGATCGCGCTAGCA 473
Qy 131 LysThrLysAsp 134
Db 472 CTGCTCAAAAGAT 461

RESULT 40
LOCUS      BH161951/c
DEFINITION ENTFRB12TR Entamoeba histolytica Sheared DNA Entamoeba histolytica
genomic, DNA sequence.
ACCESSION  BH161951
VERSION     BH161951.1  GI:15735389
KEYWORDS   GSS.
SOURCE     Entamoeba histolytica.
ORGANISM   Entamoeba histolytica.
REFERENCE  1 (bases 1 to 868)
AUTHORS    Loftus,B., Wang,Z., Van Aken,S. and Fraser,C.
TITLE      Determination of clone end sequences from Entamoeba histolytica
HMI:IMSS sheared DNA library (2001)
JOURNAL    Unpublished (2001)
COMMENT    Contact: Brendan J Loftus
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0208
            Fax: 301 838 3543
            Email: bjloftus@tigr.org
            Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
            DNA library
            Seq primer: M13-Reverse
            Class: shotgun
            High quality sequence start: 13
            High quality sequence stop: 861.
            location/Qualifiers
                1..868
                    /organism="Entamoeba histolytica"
                    /strain="HMI:IMSS"
                    /db_xref="taxon:5759"
                    /clone_lib="Entamoeba histolytica Sheared DNA"
                    /note="Vector: pHO51; Site_1: Bst I; Constructed at The
                    Institute for Genomic Research (TIGR), Rockville, MD.

```



Genomic DNA isolated from broth cultures of *E. histolytica* using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) *Entamoeba histolytica*: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Borell, Oxford University Press, 1999).

BASE COUNT 97 a 200 c 99 g 472 t  
ORIGIN

Alignment Scores:  
Pred. No.: 0.907 Length: 868  
Score: 101.50 Matches: 38  
Percent Similarity: 50.82% Conservative: 24  
Best Local Similarity: 31.15% Mismatches: 51  
Query Match: 13.06% Indels: 9  
DB: 12 Gaps: 4

US-09-847-539A-6 (1-159) x BH161951 (1-868)

Qy	33	GlulGluArgAlaIleAspGluLeuLysGlnAlaIleGluAspLysGluAlaThrThr	52
Db	610	GAAGCAAAAGCCAAAGAGAGCTGAAGCAAAAGCTAAAGAGAGAGCTGAAGCAAAAGCT	551
Qy	53	AlaIleGluAlaAlaSerSerAspAlaLeuGluAlaLeuAla-----AspGlnThr	69
Db	550	AAAGCAAGAGCTGAAGCAAAAGCTAAAGAGAGAGCTGAAGCAAAAGCTAAAGAGAGAGCT	491
Qy	70	AspAlaLeuGlnSerGluGluAla-----AlaValValLysAlaAspAsnAlaAla	86
Db	490	GAAGCAAAAGCCAAAGAGAGAGCTGAAGCAAAAGCTAAAGAGAGAGCTGAAGCAAAAGCC	431
Qy	87	SerAspAlaLeuGluAlaLeuAla---AspGlnThrAspAlaLeuGlnSerGluGluAla	105
Db	430	AAAGCAAGAGCTGAAGCAAAAGCTAAAGAGAGAGCTGAAGCAAAAGCTAAAGAGAGAGCT	371
Qy	106	GluValValGlnSerAspAsnAlaAlaSerAspAlaTrpGluLysAlaAlaThrProIle	125
Db	370	GAAGCAAAAGCTAAAGAGAGAGCTGAAGCAAAAGCTAAAGAGAGAGCTGAAGCAAAAGCT	311
Qy	126	AlaLeuAspValLysThrLysAspThrLysProValValLysLysGluGluArgGln	145
Db	310	AAAGCAAGAGTTAA-----GAAGAAATAAGAGAGAGTTAAAGAGAGTTAAAGAA	257
Qy	146	AsnVal	147
Db	256	GAAGTT	251

Search completed: October 13, 2002, 06:10:13  
Job time : 1886.39 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2002, 03:46:02 ; Search time 68,8717 Seconds  
(without alignments)  
399.383 Million cell updates/sec

Title: US-09-847-539A-6

Perfect score: 777

Sequence: 1 VDSPIEQRIIPNGTGLNL.....KKEERQNVNLTPTTGEESNP 159

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_19:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_rvirus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	777	100.0	217	16	Q9X5C4
2	760	97.8	156	2	Q9S6G4
3	754	97.0	156	2	Q9S6G3
4	694	89.3	268	2	Q9X5C6
5	692	89.1	239	2	Q9X5C5
6	151.5	19.5	439	2	Q5105
7	142	18.3	103	2	Q54180
8	138.5	17.8	323	2	Q53900
9	122.5	15.8	1110	13	Q91255
10	117	15.1	7659	5	Q94M4
11	116	14.9	166	10	Q9SCK5
12	114.5	14.7	413	2	Q53974
13	114.5	14.7	850	2	Q935B1
14	112.5	14.5	425	2	Q54744
15	111	14.3	388	2	Q9AMM4
16	111	14.3	415	2	Q54859

17 111 14.3 619 5 Q95S18  
18 111 14.3 2055 2 O85472  
19 110 14.2 890 2 Q01891  
20 109 14.0 316 10 Q9LK53  
21 109 14.0 425 2 Q54860  
22 108 13.9 159 10 Q23881  
23 108 13.9 405 2 Q54862  
24 108 13.9 415 2 P95813  
25 108 13.9 425 2 P95808  
26 108 13.9 655 2 Q9R1F5  
27 108 13.9 679 2 Q9Z4H8  
28 108 13.9 1096 5 Q24478  
29 108 13.9 1096 5 Q9VFAL  
30 107 13.8 384 2 P95810  
31 107 13.8 388 2 Q53474  
32 106.5 13.7 437 2 Q9EYB8  
33 105.5 13.6 314 2 P71213  
34 105.5 13.6 389 2 Q9Z1X5  
35 105 13.5 387 2 Q54842  
36 105 13.5 524 2 Q9K2J9  
37 105 13.5 664 2 Q53975  
38 104.5 13.4 843 2 Q47802  
39 104.5 13.4 924 5 Q9BHC4  
40 104.5 13.4 1822 2 Q07290  
41 104 13.4 576 2 Q9KJJ3  
42 104 13.4 891 2 Q04111  
43 104 13.4 1795 2 Q9LCJ9  
44 104 13.4 2478 2 Q9RL69  
45 104 13.4 2481 16 Q99QR6

#### ALIGNMENTS

#### RESULT 1

Q9X5C4 PRELIMINARY; PRT; 217 AA.  
ID Q9X5C4  
AC Q9X5C4  
DT 01-NOV-1999 (TREMBLrel. 12, Created)  
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE GRAB PRECURSOR (PROTEIN GRAB) (PROTEIN G-RELATED ALPHA 2M-BINDING DE PROTEIN).  
GN GRAB OR SPV1357.  
OS Streptococcus pyogenes.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1314;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC700294;  
RA Rasmussen M., Muller H.P., Bjorck L.;  
RT "protein GRAB of Streptococcus pyogenes regulates proteolysis at the bacterial surface by binding alpha2-macroglobulin."  
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-SF370 / ATCC 700294 / SEROTYPE M1;  
RX MEDLINE=21192684; PubMed=11296296;  
RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K., Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P., Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J., Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;  
RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes.";  
RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).  
DR EMBL: AF124399; AAD26338.1; -  
DR EMBL: AE006573; AAK34185.1; -  
DR InterPro: IPR001899; Gram\_pos\_anchor.  
DR Pfam: PF00746; Gram\_pos\_anchor; 1.  
DR PRINTS: PR00015; GPOSANCHOR.  
DR PROSITE: PS00343; GRAM\_POS\_ANCHORING; UNKNOWN\_1.  
KW Complete proteome; signal; Transmembrane.  
FT SIGNAL 1 33 POTENTIAL.

```

FT CHAIN 34 217 GRAB.
SQ SEQUENCE 217 AA; 22836 MW; 79AA8C4FF5F3FA06 CRC64;

Query Match 100.0%; Score 777; DB 16; Length 217;
Best Local Similarity 100.0%; Pred. No. 2.3e-43;
Matches 159; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VDSPIEQPRIIPNGGTLNLLGNAPEKALRNEERAIDELKKQATIEDKEATTATEAASSD 60
Db 34 VDSPIEQPRIIPNGGTLNLLGNAPEKALRNEERAIDELKKQATIEDKEATTATEAASSD 93

Qy 61 ALEALADQTDALQSEEAHVKADNAASDALEALADQTDALQSEEAHVQSDNAASDAWEK 120
Db 94 ALEALADQTDALQSEEAHVKADNAASDALEALADQTDALQSEEAHVQSDNAASDAWEK 153

Qy 121 AATPIALDVKKTKDTPVVKKEERQNVNTLPTTGESNP 159
Db 154 AATPIALDVKKTKDTPVVKKEERQNVNTLPTTGESNP 192

RESULT 2
Q9S6G4 PRELIMINARY; PRT; 156 AA.
AC Q9S6G4;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE GRAB (FRAGMENT).
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99269061; PubMed=10336419;
RA Rasmussen M., Muller H.P., Bjorck L.;
RT "Protein GRAB of streptococcus pyogenes regulates proteolysis at the
bacterial surface by binding alpha2-macroglobulin.";
RL J. Biol. Chem. 274:15336-15344(1999).
DR EMBL; AF124401; AAD26340.1; -.
DR InterPro; IPR001899; Gram_pos_anchor.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
FT NON_TER 1 156
FT NON_TER 156 156
SQ SEQUENCE 156 AA; 16575 MW; 0C20967E2FB0D866 CRC64;

Query Match 97.8%; Score 760; DB 2; Length 156;
Best Local Similarity 100.0%; Pred. No. 2e-42;
Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VDSPIEQPRIIPNGGTLNLLGNAPEKALRNEERAIDELKKQATIEDKEATTATEAASSD 60
Db 1 VDSPIEQPRIIPNGGTLNLLGNAPEKALRNEERAIDELKKQATIEDKEATTATEAASSD 60

Qy 61 ALEALADQTDALQSEEAHVKADNAASDALEALADQTDALQSEEAHVQSDNAASDAWEK 120
Db 61 ALEALADQTDALQSEEAHVKADNAASDALEALADQTDALQSEEAHVQSDNAASDAWEK 120

Qy 121 AATPIALDVKKTKDTPVVKKEERQNVNTLPTTGEE 156
Db 121 AATPIALDVKKTKDTPVVKKEERQNVNTLPTTGEE 156

RESULT 3
Q9S6G3 PRELIMINARY; PRT; 156 AA.
AC Q9S6G3;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE GRAB (FRAGMENT).
OS Streptococcus pyogenes.

```

```

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99269061; PubMed=10336419;
RA Rasmussen M., Muller H.P., Bjorck L.;
RT "Protein GRAB of streptococcus pyogenes regulates proteolysis at the
bacterial surface by binding alpha2-macroglobulin.";
RL J. Biol. Chem. 274:15336-15344(1999).
DR EMBL; AF124401; AAD26339.1; -.
DR InterPro; IPR001899; Gram_pos_anchor.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
FT NON_TER 1 156
FT NON_TER 156 156
SQ SEQUENCE 156 AA; 16587 MW; 19FAA9800599D866 CRC64;

Query Match 97.0%; Score 754; DB 2; Length 156;
Best Local Similarity 99.4%; Pred. No. 5e-42;
Matches 155; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 VDSPIEQPRIIPNGGTLNLLGNAPEKALRNEERAIDELKKQATIEDKEATTATEAASSD 60
Db 1 VDSPIEQPRIIPNGGTLNLLGNAPEKALRNEERAIDELKKQATIEDKEATTATEAASSD 60

Qy 61 ALEALADQTDALQSEEAHVKADNAASDALEALADQTDALQSEEAHVQSDNAASDAWEK 120
Db 61 ALEALADQTDALQSEEAHVKADNAASDALEALADQTDALQSEEAHVQSDNAASDAWEK 120

Qy 121 AATPIALDVKKTKDTPVVKKEERQNVNTLPTTGEE 156
Db 121 AATPIALDVKKTKDTPVVKKEERQNVNTLPTTGEE 156

RESULT 4
Q9X5C6 PRELIMINARY; PRT; 268 AA.
AC Q9X5C6;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE GRAB (FRAGMENT).
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=AP49;
RX MEDLINE=99269061; PubMed=10336419;
RA Rasmussen M., Muller H.P., Bjorck L.;
RT "Protein GRAB of streptococcus pyogenes regulates proteolysis at the
bacterial surface by binding alpha2-macroglobulin.";
RL J. Biol. Chem. 274:15336-15344(1999).
DR EMBL; AF124403; AAD26342.1; -.
DR InterPro; IPR001899; Gram_pos_anchor.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
FT NON_TER 1 268
FT NON_TER 268 268
SQ SEQUENCE 268 AA; 27744 MW; 1C59239260CDC7E7 CRC64;

Query Match 89.3%; Score 694; DB 2; Length 268;
Best Local Similarity 58.2%; Pred. No. 7.1e-38;
Matches 156; Conservative 0; Mismatches 0; Indels 112; Gaps 1;

Qy 1 VDSPIEQPRIIPNGGTLNLLGNAPEKALRNEERAIDELKKQATIEDKEATTATEAAS-- 58
Db 1 VDSPIEQPRIIPNGGTLNLLGNAPEKALRNEERAIDELKKQATIEDKEATTATEAASSD 60

Qy 59 -----
Db 61 ALEALADQTDALQSEEAHVQSDNAASDALEALADQADALQSEEAHVQSDNAAGDALEA 120

```

```
Qy 59 -----SDALEALADQ 68
Db 121 LADQTDALQSEAAVVKADNAASDALEALADQTDALQSEAAVVKADNAASDALEALADQ 180
Qy 69 TDALQSEAAVVKADNAASDALEALADQTDALQSEAAVVKADNAASDALEALADQ 128
Db 181 TDALQSEAAVVKADNAASDALEALADQTDALQSEAAVVKADNAASDALEALADQ 240
Qy 129 VKTKDTPVVKKEERQNVNLTPTTGE 156
Db 241 VKTKDTPVVKKEERQNVNLTPTTGE 268

RESULT 5
Q9X5C5
ID Q9X5C5 PRELIMINARY; PRT; 239 AA.
AC Q9X5C5;
DT 01-NOV-1999 (Tremblrel. 12, Created)
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE GRAB (FRAGMENT)
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
NCBI_TaxID=1314;
RN 1
RP SEQUENCE FROM N.A.
RC STRAIN=KTL9;
RX MEDLINE=99269061; PubMed=10336419;
RA Rasmussen M., Muller H.P., Bjorck L.;
RT "Protein GRAB of streptococcus pyogenes regulates proteolysis at the
RT bacterial surface by binding alpha2-macroglobulin.";
RL J. Biol. Chem. 274:15336-15344(1999).
DR EMBL; AF124402; AAD26341.1; -.
DR InterPro; IPR001899; Gram_pos_anchor.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
FT NON_TER 1
FT NON_TER 239
FT SEQUENCE 239 AA; 24846 MW; 61AC4F6F863AF0F5 CRC64;

Query Match 89.1%; Score 692; DB 2; Length 239;
Best Local Similarity 64.0%; Pred. No. 8.4e-38;
Matches 153; Conservative 0; Mismatches 2; Indels 84; Gaps 1;

Qy 1 VDSPIEQPRIIPNGGTLNLLGNAPKALNEERAIDELKKQAEIKLKESEKAEATQTPKEPVP 60
Db 1 VDSPIEQPRIIPNGGTLNLLGNAPKALNEERAIDELKKQAEIKLKESEKAEATQTPKEPVP 60
Qy 61 A-----
Db 61 ALEALADQADALQSEAAVVKADNAASDALEALADQTDALQSEAAVVKADNAASDTLEA 120
Qy 62 -----LEALADQTDALQSEAAVVKADNAASDALEALADQ 96
Db 121 LADQTDALQSEAAVVKADNAASDTLEALADQTDALQSEAAVVKADNAASDTLEALADQ 180
Qy 97 TDALQSEAAVVKADNAASDALEALADQTDALQSEAAVVKADNAASDTLEALADQ 155
Db 181 TDALQSEAAVVKADNAASDALEALADQTDALQSEAAVVKADNAASDTLEALADQ 239

RESULT 6
Q95105
ID Q95105 PRELIMINARY; PRT; 439 AA.
AC Q95105;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE MULTIPLE LIGAND-BINDING PROTEIN 1 PRECURSOR.
OS Streptococcus sp.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
```

```
OC Streptococcus.
NCBI_TaxID=1306;
RN 1
RP SEQUENCE FROM N.A.
RC STRAIN=SEROTYPE C;
RX MEDLINE=96202013; PubMed=8615832;
RA Talay S.R., Grammel M.P., Chhatwal G.S.;
RT "Structure of a group C streptococcal protein that binds to
RT fibrinogen, albumin and immunoglobulin G via overlapping modules.";
RL Biochem. J. 315:577-582(1996).
DR EMBL; X84989; CAA59349.1; -.
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR003345; M_repeat.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR Pfam; PF02370; M; 6.
DR PRINTS; PR00015; GPOSANCHOR.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
KW Signal; Transmembrane.
FT SIGNAL 1
FT SIGNAL 36
FT SEQUENCE 439 AA; 46527 MW; D0AE6C2A85D46B89 CRC64;

Query Match 19.5%; Score 151.5; DB 2; Length 439;
Best Local Similarity 27.2%; Pred. No. 0.018;
Matches 65; Conservative 25; Mismatches 46; Indels 103; Gaps 10;

Qy 19 NLGNA-----PEKALRNEERAIDELKK-----QATDEKEATTAIEAA 57
Db 181 NLGNAKDQLDKLTTEKESLSKDEALDERKNVLEASRTNRDLEAARDAKKAETEAEALA 240
Qy 58 SSDA-----LEALADQTDALQSEAAVVKAD-----DALQSEAAVVKAD----- 83
Db 241 ETNAKVDKLEEKQVLEASRTNRDLEAARDAKKAETEAEALAKNELAQILEASRTNR 300
Qy 84 --NAASDALEA-----LADQTDALQSEAAVVKADNAASDAWEKA----- 121
Db 301 DLEAARDAKKAVDAELAKLKAEEALKEQALKAQAEIKLKESEKAEATQTPKEPVP 360
Qy 122 -----ATPIALDVKKTKDTPV-----VKKEERQNVNLTPTTGESNP 159
Db 361 GKPSMPWTGLTPATPIAKRRKT-DVKPAKANKMVPDVKDKERK-----LPSTGETVNP 414

RESULT 7
Q54180
ID Q54180 PRELIMINARY; PRT; 103 AA.
AC Q54180;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE PROTEIN G'.
GN PROTEIN G' GENE.
OS Streptococcus sp. 'group G'.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
NCBI_TaxID=1320;
RN 1
RP SEQUENCE FROM N.A.
RC STRAIN=G148;
RX MEDLINE=90226312; PubMed=2183792;
RA Goward C.R., Murphy J.P., Atkinson T., Barstow D.A.;
RT "Expression and purification of a truncated recombinant streptococcal
RT protein G.";
RL Biochem. J. 267:171-177(1990).
DR EMBL; X53324; CAA37409.1; -.
DR SEQUENCE 103 AA; 10828 MW; E7AE8BD1C073423C CRC64;

Query Match 18.3%; Score 142; DB 2; Length 103;
Best Local Similarity 57.9%; Pred. No. 0.014;
Matches 33; Conservative 3; Mismatches 21; Indels 0; Gaps 0;

Qy 1 VDSPIEQPRIIPNGGTLNLLGNAPKALNEERAIDELKKQAEIKLKESEKAEATQTPKEPVP 57
Db 34 VDSPIEQPRIIPNGGTLNLLGNAPKALNEERAIDELKKQAEIKLKESEKAEATQTPKEPVP 90
```

```
RESULT 8
Q53900
ID Q53900 PRELIMINARY; PRT; 323 AA.
AC Q53900;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE ALBUMIN-BINDING PROTEIN (FRAGMENT).
OS Streptococcus canis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DG12;
RX MEDLINE=92363555; PubMed=1500168;
RA Sjobring U.;
RT "Isolation and molecular characterization of a novel albumin-binding
RL protein from group G streptococci.";
DR EMBL; M95520; AAA26847.1; -
DR HSSP; Q51911; IGAB.
DR InterPro; IPR002988; GA.
DR InterPro; IPR001899; Gram_pos_anchor.
DR Pfam; PF01468; GA; 2.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
FT NON_TER 1
FT NON_TER 323
FT SEQUENCE 323 AA; 34990 MW; E6903CFDCD5D373E CRC64;

Query Match 17.8%; Score 138.5; DB 2; Length 323;
Best Local Similarity 33.9%; Pred. No. 0.088;
Matches 39; Conservative 25; Mismatches 36; Indels 15; Gaps 4;

QY 42 QKAEDEKATTAIE---AASSD-----ALEALADQTDALQSEAAVYKADNAASDALEAL 93
DB 34 ESSIKDKAQLDSYKIKASVDRDEITELSAQADQIVSAQADNEAITKAEEDSSKAWEA 93
QY 94 ADQTDALQSEAEVQSDNAASDAWEKAATPIALDVKKTKDTPVVKKEERQNVN 148
DB 94 ADQANTAKAEDELAKAEKSSDAWEKAA---ALDQAK----QAALKEFDRYGVS 141

RESULT 9
Q91255
ID Q91255 PRELIMINARY; PRT; 1110 AA.
AC Q91255;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE NF-180.
OS Petromyzon marinus (Sea lamprey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Hyperoartia;
OC Petromyzontiformes; Petromyzontidae; Petromyzon.
OX NCBI_TaxID=7757;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=CNS;
RX MEDLINE=95287814; PubMed=7770000;
RA Jacobs A.J., Kamholz J., Selzer M.E.;
RT "The single lamprey neurofilament subunit (NF-180) lacks
RT multiphosphorylation repeats and is expressed selectively in
RT projection neurons.";
RL Brain Res. Mol. Brain Res. 29:43-52(1995).
DR EMBL; U19361; AAA80106.1; -
DR InterPro; IPR001664; IF.
DR Pfam; PF00038; filament; 1.
DR PRINTS; PR00194; TROPOMYOSIN.
DR PROSITE; PS00226; IF; UNKNOWN_1.
FT SEQUENCE 1110 AA; 123818 MW; 6558DA73DAF6974C CRC64;
```

```
Query Match 15.8%; Score 122.5; DB 13; Length 1110;
Best Local Similarity 29.4%; Pred. No. 3.8;
Matches 37; Conservative 28; Mismatches 50; Indels 11; Gaps 5;

QY 38 DELKQKQAIEDKEATTAIEAASSDALEALAD-QTDALQSEAAVYKADNAASDALEALA 94
DB 705 DEAESEEVKEEVT-----KSDAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEV 758
QY 95 DQTDALQSE-EAEVQSDNAASDAWEKAATPIALDVKKTKDTPVVKKEERQNVNLTPT 153
DB 759 EETEAATEAEAEASDDEKPEEVEKSEAPVAPEAKKAPKAPKAPKAKVES-PTS 817
QY 154 GEBSNP 159
DB 818 EPEDEP 823

RESULT 10
Q9N4M4
ID Q9N4M4 PRELIMINARY; PRT; 7659 AA.
AC Q9N4M4;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE HYPOTHETICAL 854.0 KDA PROTEIN.
GN ZK973.6.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=99069613; PubMed=98519116;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Gattung S., Goela D., Broly M.;
RT "The sequence of C. elegans cosmid ZK973.";
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Waterston R.;
RT "Direct Submission.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC006834; AAF40010.2; -
KW Hypothetical protein.
SQ SEQUENCE 7659 AA; 854006 MW; BBA95964D79B4FF1 CRC64;

Query Match 15.1%; Score 117; DB 5; Length 7659;
Best Local Similarity 24.9%; Pred. No. 76;
Matches 49; Conservative 28; Mismatches 66; Indels 54; Gaps 9;

QY 9 RIIPNGGTLNLLGNA-----PEKLALR-----NEERADIELKQA- 44
DB 6812 RIAPHILATLVEAVNDVPASVEPSAVALRDRAAKFVSDLEKNIQKTGDEKRADELKNDVG 6871
QY 45 -----TED-----KEATTAEASSDA-----LEALADQTDALQSEAAVYKADNA 85
DB 6872 NAVKNVEDVYVYQNPQPPQDLVAKDDANKLKATVEQLTKLAESSDKIDVQVAKIDKSKT 6931
QY 86 AS-----DALEALADQTDALQSEAEVQSDNAASDAWEKAATPIALDVKKTKDTPVVK 141
DB 6932 KAKELLOALEKAIPQEDAIRRQAEI-----NDRLNKLEKELTKV--DEPKPEDALPIV-D 6984
QY 142 EERQNVNLTPTTGEEN 158
```

Db 6985 QLAANTNTLKTATDSNN 7001

# RESULT 11

Q93CK5 PRELIMINARY; PRT; 166 AA.  
 AC Q93CK5; 14.9%; Score 116; DB 10; Length 166;  
 DT 01-MAY-2000 (Tremblrel. 13, Created)  
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
 DE HYPOTHETICAL 17.4 KDA PROTEIN.  
 GN T9C5.130.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Rieger M., Gabel C., Mueller-Auer S., Schaefer M., Zipp M.,  
 RA Mewes H.W., Lemcke K., Mayer K.F.X., Quetier F., Salancoubat M.;  
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA EU Arabidopsis sequencing project;  
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AL132964; CAB62458.1; -  
 KW Hypothetical protein.  
 SQ SEQUENCE 166 AA; 17373 MW; 3C612A36544F57FE CRC64;

Query Match 14.9%; Score 116; DB 10; Length 166;  
 Best Local Similarity 26.3%; Pred. No. 1-2;  
 Matches 41; Conservative 25; Mismatches 62; Indels 28; Gaps 5;  
 Qy 12 PNGTTLNLGNAP-----EKLALNREERAIDELKQAIEDKEATTATAEAASSDALEALA 66  
 Db 8 PKESDMONEGSPVKNPVSENVVAKENNTGESGKQNTVAETTTSVEAKETPPVEPTK 67  
 Qy 67 DQDALQSE-----EAAVV---KADNAASDALEALADQTDALQSEAEVQVS 110  
 Db 68 ETTPAVQEVAAVSESSADAGEAAVPAKVENAENAEKVAEVAVAAPKEV--- 124  
 Qy 111 DNAASDAWEKA-ATPIALDVKKTKDTPVVKKEERQ 145  
 Db 125 ---AVEAEKKAEPVKAEPVKAEPVKEESKQ 157

# RESULT 12

Q53974 PRELIMINARY; PRT; 413 AA.  
 AC Q53974;  
 DT 01-NOV-1996 (Tremblrel. 01, Created)  
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
 DE MAG PROTEIN PRECURSOR.  
 GN MAG.  
 OS Streptococcus dysgalactiae  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1334;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94259307; PubMed=7515368;  
 RA Jonsson H., Frykberg L., Rantamaeki L., Guss B.;  
 RT "MAG, a novel plasma protein receptor from Streptococcus  
 RT dysgalactiae";  
 RL Gene 143:85-89(1994).  
 DR EMBL; L27798; AAA26921.1; -  
 DR HSP; P06654; IPCX  
 DR InterPro; IPR002988; GA.  
 DR InterPro; IPR001899; Gram\_pos\_anchor.  
 DR InterPro; IPR000724; Igg\_bind\_B.  
 DR Pfam; PF01468; GA; 1.

DR Pfam; PF00746; Gram\_pos\_anchor; 1.  
 DR Pfam; PF01378; Igg\_binding\_B; 1.  
 DR PRINTS; PR00015; GPANCHOR.  
 DR PROSITE; PS00343; GRAM\_POS\_ANCHORING; UNKNOWN\_1.  
 KW Signal; Transmembrane.  
 FT SIGNAL 1 34 POTENTIAL.  
 FT CHAIN 35 413 POTENTIAL.  
 SQ SEQUENCE 413 AA; 44004 MW; F04DC71044F9E50F CRC64;

Query Match 14.7%; Score 114.5; DB 2; Length 413;  
 Best Local Similarity 26.2%; Pred. No. 4.2;  
 Matches 49; Conservative 15; Mismatches 56; Indels 67; Gaps 6;

Qy 36 AIDELKKQATD-----KEATTATD---AASSDALEALADQTDALQSEAEAVVK 81  
 Db 206 AIELKKYGGDYIKLINCKTAEGVTALKDEILASKPAVIDAPELTPALTYKLVI-- 263  
 Qy 82 ADNAASDALEALADQTDALQSEEA-EVVSQSDNAASDAWE----- 119  
 Db 264 --NGKTLKGETTTKAVDAETAETAKAFKOVANENGVDGVVYDDATKTFVTETVPGDA 321  
 Qy 120 -----KAATPIALDVKKTKDT-----KPVVKKEERQNVNLTPT 152  
 Db 322 PTEPKKPEASTPLVPLTPATPIAKDDAKDDTKKDDAKKPEAKKEEAKKAATLPT 381  
 Qy 153 TGEESNP 159  
 Db 382 TGEESNP 388

# RESULT 13

Q935B1 PRELIMINARY; PRT; 850 AA.  
 AC Q935B1;  
 DT 01-DEC-2001 (Tremblrel. 19, Created)  
 DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)  
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
 DE PUTATIVE PHAGE TAIL PROTEIN.  
 GN HCM2.0051C.  
 OS Salmonella enterica subsp. enterica serovar Typhi.  
 OG Plasmid phCw2.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Salmonella.  
 OX NCBI\_TaxID=90370;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CT18;  
 RX MEDLINE=21534947; PubMed=11677608;  
 RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,  
 RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,  
 RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,  
 RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,  
 RA Felwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,  
 RA Krog A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,  
 RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,  
 RA Whitehead S., Barrell B.G.;  
 RT "Complete genome sequence of a multiple drug resistant Salmonella  
 RT enterica serovar typhi CT18";  
 RL Nature 413:848-852(2001).  
 DR EMBL; AL513384; CAD09918.1; -  
 KW Plasmid.  
 SQ SEQUENCE 850 AA; 90917 MW; 21271493A07A6563 CRC64;

Query Match 14.7%; Score 114.5; DB 2; Length 850;  
 Best Local Similarity 31.0%; Pred. No. 9.4;  
 Matches 40; Conservative 20; Mismatches 44; Indels 25; Gaps 5;

Qy 37 IDELKKQAIEDKEATTATAEAASSDAL-----EALADQTDALQSEAEAVVKADNA 85  
 Db 146 VDTIESSSKALAAARSEAAAEALNLSKQAAAGSEANAKASENNAASQQAATSESA 205  
 Qy 86 ASDALEALADQTDALQSEAEVVSQSDNAASDAWEKAATPIALDVKK---TKDTPVVKKE 142





```
Db 275 ANAEKELQSOAALEKQLEATKRELADLOAKLAATNQEKLEAEAKALKQOL-AKQVE 333
QY 104 EAEVQSDNAAADAWAKAATPIALDVKTKDTRPVVKKBERQNVN-----TLPTTG 154
Db 334 ELAKLKADKASG-----AQKP---DTRKPDNKEVTRPSQTRTNTNKAPMPQTKRQLPSTG 385
QY 155 EE-SNP 159
Db 386 EETNP 391

RESULT 17
Q95S18 PRELIMINARY; PRT; 619 AA.
AC Q95S18; 2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DE 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE HU01392P.
GN FAU.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Y, CN BW SP;
RA Scapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nunoo J., Pacle J., Paragas V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celnikier S.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY060997; AAL28545.1; -.
SQ SEQUENCE 619 AA; 68641 MW; 7DCB26AD1AB9CEE CRC64;

Query Match 14.3%; Score 111; DB 5; Length 619;
Best Local Similarity 31.5%; Pred. No. 11;
Matches 41; Conservative 20; Mismatches 43; Indels 26; Gaps 6;

QY 4 PIEQRIIPNGGTLNLLGNAPEKALRNE-----EPAIDELKKQAIEDKEATTAEEA 57
Db 297 PLEPLDV-----EAKKQRLQERLLTVNEEALDEVD---LEKKRAQKADEAK 342
QY 58 SSDALEALADOTDAL--QSEAAVVKADNAAADALEALADQ---TDALQSEEAEEVQSDN 112
Db 343 RREE-RALKERDRLEAEKQAAKAAKAAEAAKTAABEALLAEAAQAAEAEAKALK 401
QY 113 AASDAWEKAA 122
Db 402 AAEDAQKAA 411

RESULT 18
O85472 PRELIMINARY; PRT; 2055 AA.
AC O85472; 1998 (Tremblrel. 08, Created)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE EXTRACELLULAR MATRIX BINDING PROTEIN (FRAGMENT).
GN EMB.
OS Abiotrophia defectiva.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Abiotrophia.
OX NCBI_TaxID=46125;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NVS-47;
RX MEDLINE=99081722; PubMed=9864195;
RA Manganello R., van de Rijn I.;
RT *Characterization of emb, a gene encoding the major adhesin of
```

```
RT Streptococcus defectivus.*;
RL Infect. Immun. 67:50-56(1999).
DR ENBL; AF067776; AAD03320.1; -.
FT NON_TER 2055 2055
SQ SEQUENCE 2055 AA; 215642 MW; 9699C11DD93E2FD CRC64;

Query Match 14.3%; Score 111; DB 2; Length 2055;
Best Local Similarity 25.6%; Pred. No. 43;
Matches 43; Conservative 28; Mismatches 61; Indels 36; Gaps 6;

QY 18 TNLIGNAPEKALRNEERAIDE----LKQAIED-----KEATTA 53
Db 1241 TQAINNVQTPAAKNAKAAVEQAADAKQAIEPNLTROEKDAATAKVQDQETNKARQA 1300
QY 54 IEAASSDALEALADOTDALQSEEA-----AVVKADNAAADALEALADOTDALQSEEA 107
Db 1301 IDAATTNA-DVTAKONEGTQAINAVPQTPKAKTDAKNAVQAEE---DKSAIENDPNLT 1356
QY 108 VQSDNAAADAWEKAATPI--ALDVKTKDTPVVKKEERQNVNTLPTT 153
Db 1357 REEKDAKAKVDAEATKAKNAIDAATSNDDETAKONEGTQAINAVPQT 1404

RESULT 19
Q01891 PRELIMINARY; PRT; 890 AA.
AC Q01891;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE SURFACE EXCLUSION PROTEIN SEAL PRECURSOR.
GN SEAL.
OS Enterococcus faecalis (Streptococcus faecalis).
OG Plasmid pAD1.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Enterococcaceae;
OC Enterococcus.
OX NCBI_TaxID=1351;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-DS16;
RX MEDLINE=92293110; PubMed=1603060;
RA Weidlich G., Wirth R., Galli D.;
RT *Sex pheromone plasmid pAD1-encoded surface exclusion protein of
RT Enterococcus faecalis.*;
RL Mol. Gen. Genet. 233:161-168(1992).
DR EMBL; X62658; CAA44526.1; -.
DR PIR; S22452; S22452.
DR InterPro; IPR001899; Gram_pos_anchor.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
KW Plasmid; Signal; Membrane.
FT SIGNAL 1 27 POTENTIAL.
FT CHAIN 28 890 SURFACE EXCLUSION PROTEIN SEAL.
SQ SEQUENCE 890 AA; 98227 MW; DCEDC2F131F634FB CRC64;

Query Match 14.2%; Score 110; DB 2; Length 890;
Best Local Similarity 29.1%; Pred. No. 19;
Matches 44; Conservative 24; Mismatches 71; Indels 12; Gaps 6;

QY 7 QPRIIPNGGTLNLLGNAPEKALRNEERAIDEKQAIEDKEATTAEEAASSDALEALA 66
Db 32 QPKTPENSST-----EQPTVKATQTTTEQAITE-KQQQVTEKQIVDQKQVADTAKKEK 84
QY 67 DOTDALQSEEAAYVKADNAAASD-ALEALADOTDALQSEEAEEVQSDNAAADAWEKAATPI 125
Db 85 DAIDQSVKDDQAAVVDQNKDALDQSQAVTDQ-QAVVDEAKKV--DEATPSAIEKAKEQV 141
QY 126 ALDVKTKDTPVVKKEERQNVNTLPTTGE 156
Db 142 ATDQAVDDQOKVV-DQATDVTNQQQAVVEE 171

RESULT 20
```

2

OS Streptococcus pyogenes.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus  
OX NCBI\_TaxID=1314;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-M TYPE 76;  
RX MEDLINE=89282846; PubMed=2660147;  
RA Heath D.G., Cleary P.P.;  
RT "Fc-receptor and M-protein genes of group A Streptococci are products  
of gene duplication.";  
RL Proc. Natl. Acad. Sci. U.S.A. 86:4741-4745(1989).  
DR EMBL; M22532; AAB95296.1; -;  
DR InterPro; IPR001899; Gram\_pos\_anchor.  
DR PROSITE; PS00343; GRAM\_POS\_ANCHORING; UNKNOWN\_1.  
KW Signal; Receptor.  
FT SIGNAL 1 41 POTENTIAL.  
FT CHAIN 42 >405 FC-RECEPTOR PROTEIN.  
FT NON\_TER 405  
SQ SEQUENCE 405 AA; 44299 MW; 9B73D62D515B41F3 CRC64;

Query Match 13.9%; Score 108; DB 2; Length 405;  
Best Local Similarity 27.5%; Pred. No. 11;  
Matches 53; Conservative 25; Mismatches 67; Indels 48; Gaps 10;

QY 6 EQPRIIPNGGTLTNLIGNA-----PEKALRNEERAIDEL---KKQAIED 47  
DB 218 ETAKLQSEAAATLENLLGSAKRELTDLOAKLDTATAEKAKLESQVTTLENLLGSAKRELTD 277  
QY 48 KEATTAIEAASSDALE----ALADQTDALQSEEA---AVVKADNAASDALE----ALADQ 96  
DB 278 LQAKLDAANAEEKLQSQAAALEKOLEATKKELADLQAKLAATNOEKEKEAEAKALKEQ 337  
QY 97 TDALQSEEAEEVQSDNAASDAWEKAATPIALDVKKTKDKPVVKKKEERQNVN----- 148  
DB 338 L-AKQAEELAKLKADKASG-----AQKP---DTKPGNKEVTPRPSQTRTNTNKPAPNAQTK 388  
QY 149 -TLPTTGEE-SNP 159  
DB 389 RQLPSTGEETNP 401

RESULT 24  
P95813  
ID P95813 PRELIMINARY; PRT; 415 AA.  
AC P95813;  
DT 01-MAY-1997 (TRENBLrel. 03, Created)  
DT 01-MAY-1997 (TRENBLrel. 03, Last sequence update)  
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)  
DE FC-GAMMA RECEPTOR PRECURSOR.  
GN FCRA9.  
OS Streptococcus pyogenes.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1314;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-71-683;  
RX MEDLINE=96396849; PubMed=8803952;  
RA Krebs B., Kaufhold A., Boyle M.D.P., Podbielski A.;  
RT "Different alleles of the fcrA/mrp gene of Streptococcus pyogenes  
encode M-related proteins exhibiting an identical immunoglobulin-  
binding pattern.";  
RL Med. Microbiol. Immunol. 185:39-47(1996).  
DR EMBL; X78483; CAA55229.1; -;  
DR InterPro; IPR001899; Gram\_pos\_anchor.  
DR Pfam; PF00746; Gram\_pos\_anchor; 1.  
DR PRINTS; PR00015; GPOSANCHOR.  
DR PROSITE; PS00343; GRAM\_POS\_ANCHORING; UNKNOWN\_1.  
DE FC-GAMMA RECEPTOR PRECURSOR.  
GN FCRA9.  
OS Streptococcus pyogenes.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1314;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-71-683;  
RX MEDLINE=96396849; PubMed=8803952;  
RA Krebs B., Kaufhold A., Boyle M.D.P., Podbielski A.;  
RT "Different alleles of the fcrA/mrp gene of Streptococcus pyogenes  
encode M-related proteins exhibiting an identical immunoglobulin-  
binding pattern.";  
RL Med. Microbiol. Immunol. 185:39-47(1996).  
DR EMBL; X78482; CAA55229.1; -;  
DR InterPro; IPR001899; Gram\_pos\_anchor.  
DR Pfam; PF00746; Gram\_pos\_anchor; 1.  
DR PRINTS; PR00015; GPOSANCHOR.  
DR PROSITE; PS00343; GRAM\_POS\_ANCHORING; UNKNOWN\_1.  
KW Receptor; Signal; Transmembrane.  
FT SIGNAL 1 41 POTENTIAL.  
SQ SEQUENCE 415 AA; 45256 MW; B4C67D8E230E6DF CRC64;

Query Match 13.9%; Score 108; DB 2; Length 415;  
Best Local Similarity 27.5%; Pred. No. 11;  
Matches 53; Conservative 25; Mismatches 67; Indels 48; Gaps 10;

QY 6 EQPRIIPNGGTLTNLIGNA-----PEKALRNEERAIDEL---KKQAIED 47  
DB 208 ETAKLQSEAAATLENLLGSAKHELTDLOAKLDTATAEKAKLESQVTTLENLLGSAKRELTD 267  
QY 48 KEATTAIEAASSDALE----ALADQTDALQSEEA---AVVKADNAASDALE----ALADQ 96  
DB 268 LQAKLDAANAEEKLQSQAAALEKOLEATKKELADLQAKLAATNOEKEKEAEAKALKEQ 327  
QY 97 TDALQSEEAEEVQSDNAASDAWEKAATPIALDVKKTKDKPVVKKKEERQNVN----- 148  
DB 328 L-AKQAEELAKLKADKASG-----AQKP---DTKPGNKEVTPRPSQTRTNTNKPAPNAQTK 378  
QY 149 -TLPTTGEE-SNP 159  
DB 379 RQLPSTGEETNP 391

RESULT 25  
P95808  
ID P95808 PRELIMINARY; PRT; 425 AA.  
AC P95808;  
DT 01-MAY-1997 (TRENBLrel. 03, Created)  
DT 01-MAY-1997 (TRENBLrel. 03, Last sequence update)  
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)  
DE FC-GAMMA RECEPTOR PRECURSOR.  
GN FCRA13.  
OS Streptococcus pyogenes.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1314;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-71-686;  
RX MEDLINE=96396849; PubMed=8803952;  
RA Krebs B., Kaufhold A., Boyle M.D.P., Podbielski A.;  
RT "Different alleles of the fcrA/mrp gene of Streptococcus pyogenes  
encode M-related proteins exhibiting an identical immunoglobulin-  
binding pattern.";  
RL Med. Microbiol. Immunol. 185:39-47(1996).  
DR EMBL; X78483; CAA55232.1; -;  
DR InterPro; IPR001899; Gram\_pos\_anchor.  
DR Pfam; PF00746; Gram\_pos\_anchor; 1.  
DR PRINTS; PR00015; GPOSANCHOR.  
DR PROSITE; PS00343; GRAM\_POS\_ANCHORING; UNKNOWN\_1.  
KW Receptor; Signal; Transmembrane.  
FT SIGNAL 1 41 POTENTIAL.  
SQ SEQUENCE 425 AA; 46558 MW; 1DD73FB2CF76DAD8 CRC64;

Query Match 13.9%; Score 108; DB 2; Length 425;  
Best Local Similarity 27.5%; Pred. No. 11;  
Matches 53; Conservative 25; Mismatches 67; Indels 48; Gaps 10;

QY 6 EQPRIIPNGGTLTNLIGNA-----PEKALRNEERAIDEL---KKQAIED 47  
DB 218 ETAKLQSEAAATLENLLGSAKHELTDLOAKLDTATAEKAKLESQVTTLENLLGSAKRELTD 277  
QY 48 KEATTAIEAASSDALE----ALADQTDALQSEEA---AVVKADNAASDALE----ALADQ 96  
DB 278 LQAKLDAANAEEKLQSQAAALEKOLEATKKELADLQAKLAATNOEKEKEAEAKALKEQ 337  
QY 97 TDALQSEEAEEVQSDNAASDAWEKAATPIALDVKKTKDKPVVKKKEERQNVN----- 148  
DB 338 L-AKQAEELAKLKADKASG-----AQKP---DTKPGNKEVTPRPSQTRTNTNKPAPNAQTK 388  
QY 149 -TLPTTGEE-SNP 159  
DB 389 RQLPSTGEETNP 401

RP	SEQUENCE FROM N.A.
RR	Muscholl-Silberhorn A.B.;
RA	"Cloning and expression of Asa373, a novel adhesin unrelated to the
RT	other sex pheromone plasmid-encoded aggregation substances of
RT	<i>Enterococcus faecalis</i> .";
RL	Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR	EMBL; AJ132039; CAB38226.1; "
DR	InterPro; IPR001899; Gram_pos_anchor.
DR	Pfam; PF00746; Gram_pos_anchor: 1.
DR	PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
KW	Signal; Plasmid.
FT	Signal
FT	SEQUENCE 679 AA; 75614 MW; 5E5070689D92553A CRC64;
SD	POTENTIAL.
SD	37

[illegible]

RESULT 29

Q9VFA1	Q9VFA1	PRELIMINARY;	PRT: 1096 AA.
ID	Q9VFA1		
AC	Q9VFA1		
DT	01-MAY-2000 (TREMBLrel. 13, Created)		
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)		
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)		
DE	CEN190 PROTEIN		
DE	CEN190 OR CG6384.		
GN	CEN190 OR CG6384.		
OS	Drosophila melanogaster (Fruit fly).		
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;		
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;		
OC	Ephydroidea; Drosophilidae; Drosophila.		
OX	NCBI_TaxID=7227;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRADLINE=BERKELEY;		
RC	MEDLINE=20196006; PubMed=10731132;		
RA	Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,		
RA	Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,		
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,		
RA	Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,		
RA	Brandon R.C., Rogers J.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,		
RA	Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,		
RA	Abriil J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,		
RA	Balleg R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,		
RA	Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,		
RA	Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,		
RA	Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,		
RA	Cherry J.M., Cawley S., Dahlke C., Davidson L.B., Davies P.,		
RA	de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,		
RA	Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,		
RA	Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,		
RA	Foster C., Gabriellian A.E., Gary N.S., Gelbart W.M., Glasser K.,		
RA	Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,		
RA	Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,		
RA	Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,		
RA	Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,		
RA	Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,		
RA	Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,		
RA	Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,		
RA	Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,		
RA	Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,		
RA	Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclab J.M.,		
RA	Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,		
RA	Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,		
RA	Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,		
RA	Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,		
RA	Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,		
RA	Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,		
RA	Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,		
RA	Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,		
RA	Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,		
RA	Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.:		
RT	"The genome sequence of <i>Drosophila melanogaster</i> ."		
RL	Science 287:2185-2195(2000).		
DR	EMBL: AE003708; AAF55159.1		
DR	FlyBase: FBgn0000283; Cen190		
DR	InterPro: IPR000210; BTF_POZ		
DR	InterPro: IPR000886; ER_target		
DR	InterPro: IPR000822; Znf-C2H2		
DR	Pfam: PF00651; BTF: 1		
DR	Pfam: PF00096; Zf-C2H2; 3		
DR	SMART: SM00225; BTF: 1		
DR	SMART: SM00355; Znf-C2H2; 3		
DR	PROSITE: PS50097; BTF: 1		
DR	PROSITE: PS00014; ER_TARGET; UNKNOWN_1		
DR	PROSITE: PS00028; ZINC_FINGER_C2H2_1; 2		
DR	PROSITE: PS0157; ZINC_FINGER_C2H2_2; 2		
DR	DNA-binding; Metal-binding; Zinc-finger		
SW	SEQUENCE 1096 AA; 121678 MW; 5CDS5C492B948D39 CRC64		

```

Query Match          13.9%; Score 108; DB 5; Length 1096;
Best Local Similarity 21.8%; Pred. No. 33;
Matches 36; Conservative 42; Mismatches 61; Indels 26; Gaps 6;

QY 11 IPNGGTLNLLGNAPKALRNE-----BRAIDELKKQAIEDKEATTATMAAS-----SD 60
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 874 VPN-----INSPEKMDVDSEAADEKSKAEVQIIRKEALENDQEEFKEDSPIHSD 926
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
QY 61 ALEALADQTDALQSEAAVVKADNAASDALEALADQTDALQSEAEVQVS-DNAASDAWE 119
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 927 SVAELREAVTASEGEDDVHLEADNIRKELLDELIAEAEK-PDQEKDVIQSEENATTEALD 985
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
QY 120 KAAT-----PIALDVKKTKDTPKPVVKKEERQNVNTLPTTGES 157
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 986 RSVTDEDDLVPPTQVSTQEQMEIDEPAAEKAAENNEDTRTADEKEA 1030
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::

RESULT 30
P95810 PRELIMINARY; PRT; 384 AA.
ID AC P95810 PRELIMINARY; PRT; 384 AA.
AC AC P95810;
DT DT 01-MAY-1997 (Tremblrel. 03, Created)
DT DT 01-MAY-1997 (Tremblrel. 03, Last sequence update)
DT DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE DE FCRA8 PROTEIN PRECURSOR.
GN GN FCRA8.
OS OS Streptococcus pyogenes.
OC OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC OC Streptococcus.
OX OX NCBI_TaxID=1314;
RN RN [1]
RP RP SEQUENCE FROM N.A.
RC RC STRAIN=M8-4025;
RX RX MEDLINE=96396849; PubMed=8803952;
RA RA Krebs B., Kaufhold A., Boyle M.D.P., Podbielski A.;
RT RT "Different alleles of the fcrA/mrp gene of Streptococcus pyogenes
   encode M-related proteins exhibiting an identical immunoglobulin-
   binding pattern.";
RL RL Med. Microbiol. Immunol. 185:39-47(1996).
DR DR EMBL; X78484; CAA55235.1; .
DR DR InterPro; IPR001899; gram_pos_anchor.
DR DR Pfam; PF00746; gram_pos_anchor; 1.
DR DR PRINTS; PR00015; gram_pos_anchor.
DR DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
KW KW Signal; Transmembrane.
FT SIGNAL 1 41 POTENTIAL.
SQ SEQUENCE 384 AA; 41708 MW; 8DF120BGA30BAD5 CRC64;

Query Match          13.8%; Score 107; DB 2; Length 384;
Best Local Similarity 28.4%; Pred. No. 12;
Matches 52; Conservative 23; Mismatches 60; Indels 48; Gaps 10;

QY 16 TLTNLLGNA-----PEKALRNEERAIDEL---KKQAIEDKEATTATMAA 57
   || |||||:: || :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 187 TLENLGSAKRELTDLQAKLDTATAEKAKLESQVTTLENLGSAKRELTDLQAKLDAANA 246
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
QY 58 SSDALE-----ALADQTDALQSEEA---AVVKADNAASDALE-----ALADQTDALQSEEA 106
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 247 EKEKLSQAALAEKQLEATKTELADLQAKLAATNQEKLEAEAKALKEQL-AKQAEELA 305
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
QY 107 VVOSDNAASDAWEKAATPIALDVKKTKDTPKPVVKKEERQNVN-----TLPTTGBE- 156
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 306 KLKADKASG-----AQKP---DTPGKNKEVPTRESQTRTNTNKAPMAQTCKQLPSTGEET 357
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
QY 157 SNP 159
   :: ::
Db 358 TNP 360

RESULT 31
Q53474 PRELIMINARY; PRT; 388 AA.
ID ID Q53474
AC AC Q53474;

```

```

DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE MRP4.
GN Streptococcus pyogenes.
OS Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95147851; PubMed=7845360;
RA Katerov V., Schalen C., Totolian A.A.;
RT "Sequencing of genes within the vir regulon of Streptococcus pyogenes
RT type M15--an opacity factor-positive serotype with low opacity factor
RT expression."
RL Mol. Gen. Genet. 245:78-85(1994).
DR EMBL: S75411; AAB3261.1; -.
DR InterPro: IPR001899; Gram_pos_anchor.
DR Pfam: PF00746; Gram_pos_anchor; 1.
DR PRINTS: PR00015; GPOSANCHOR.
DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
KW Transmembrane.
SQ SEQUENCE 388 AA; 41957 MW; 2C7320D575A5C899 CRC64;

Query Match 13.8%; Score 107; DB 2; Length 388;
Best Local Similarity 30.4%; Pred. No. 12;
Matches 49; Conservative 18; Mismatches 55; Indels 39; Gaps 8;

QY 16 TLTNLIGNAPE-----KLALNEERAIDELKKQAIEDK-----EATTA-----TEAASSDA 68
DB 226 TLTNLIGNAPE-----KLALNEERAIDELKKQAIEDK-----EATTA-----TEAASSDA 68
QY 69 TDALQSEEAHVADNAASDALEALADQTDALQSEEAHVQSDNAASDAWEKAATPTALD 128
DB 273 LADLQAKLATNCEKEKLEAKALKEQL-AKQAEELAKLADKAGS-----AKRP---D 323
QY 129 VKYTKTKPVVKKERQNVN-----TLPTTGEE-SNP 159
DB 324 TKPGNEVPTPSQTRTNTNKAPMAQTKQLPSTGETNP 364

RESULT 32
Q9EYB8
ID Q9EYB8 PRELIMINARY; PRT; 437 AA.
AC Q9EYB8;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE TAIL FIBER PROTEIN.
GN H0004.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=O157:H7;
RC MEDLINE=20564182; PubMed=11111050;
RA Yokoyama K., Makino K., Kubota Y., Watanabe M., Kimura S.,
RA Yutsudo C.H., Kurokawa K., Ishii K., Hattori M., Abe H., Iida T.,
RA Yamamoto K., Hayashi T., Yasunaga T., Honda T., Sasakawa C.,
RA Shinagawa H.;
RT "Complete nucleotide sequence of the defective Sakai-VT1 prophage
RT carrying the shiga toxin 1 genes of the enterohemorrhagic Escherichia
RT coli O157:H7 derived from the Sakai outbreak."
RL Gene 258:127-139(2000).
DR EMBL: AP000400; BAB19560.1; -.
DR InterPro: IPR000087; Collagen.
DR Pfam: PF01391; Collagen; 2.
SQ SEQUENCE 437 AA; 44183 MW; 408C870ACD31E14D CRC64;

Query Match 13.7%; Score 106.5; DB 2; Length 437;

```

```

Best Local Similarity 28.5%; Pred. No. 15;
Matches 47; Conservative 24; Mismatches 67; Indels 27; Gaps 7;

QY 15 GTLTNLIGNAPE-----KLALNEERAIDELKKQAIEDK-----EATTA-----TEAASSDA 61
DB 84 GTLNLFGANSEDDVRPEALRFEELMVEEAARHAEAKKNAAGEAETSARNAGISASQAE 143
QY 62 LEALADQTD--DALQS-----EEAAVVK-ADNAASDALEALADQTDALQSEEAHVQSDNA 113
DB 144 SAANADTSAGDALESARQAASAAAKQSDASSSSASASAAQAASQSSQSAAEUSRKT 203
QY 114 ASDAWEKAATPTIALDVKKTKDTPVVKKEER-----QNVNTLPT 152
DB 204 AESAGNAARDATTATEKARESAESQAESQSRIAAEAVNRPT 248

RESULT 33
P71213
ID P71213 PRELIMINARY; PRT; 314 AA.
AC P71213;
DT 01-FEB-1997 (TReMBLrel. 02, Created)
DT 01-FEB-1997 (TReMBLrel. 02, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE ESPB GENE (SECRETED PROTEIN ESPB).
GN ESPB.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=STEC;
RC MEDLINE=97045129; PubMed=8890194;
RA Ebel F., Deibel C., Kresse A.U., Guzman C.A., Chakraborty T.;
RT "Temperature- and medium-dependent secretion of proteins by Shiga
RT toxin-producing Escherichia coli."
RL Infect. Immun. 64:4472-4479(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=DEC-1;
RX MEDLINE=97427930; PubMed=9284118;
RA Abe A., Kenny B., Stein M., Finlay B.B.;
RT "Characterization of two virulence proteins secreted by rabbit
RT enteropathogenic Escherichia coli, EspA and EspB, whose maximal
RT expression is sensitive to host body temperature."
RL Infect. Immun. 65:3547-3555(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=RDEC-1;
RA Abe A., Kenny B., Stein M., Finlay B.B.;
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 8-308 FROM N.A.
RC STRAIN=85/150;
RA Abe A., Heczek U., Heggele R.G., Finlay B.B.;
RT "Two Enteropathogenic Escherichia coli Type III Secreted Proteins,
RT EspA and EspB, are Virulence Factors."
RL J. Exp. Med. 0:0-0(1998).
RN [5]
RP SEQUENCE OF 8-308 FROM N.A.
RC STRAIN=85/150;
RA Lau G., Abe A., Finlay B.B.;
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=B10;
RA Nougayrede J.P., Marches O., Boury M., Mainil J., Charlier G.,
RA Pohl P., De Rycke J., Milon A., Oswald E.;
RT "The long-term cytoskeletal rearrangement induced by rabbit
RT enteropathogenic Escherichia coli is Esp dependent but intimin
RT independent."
RL Mol. Microbiol. 31:19-30(1999).

```

```
[7]
RP SEQUENCE FROM N.A.
RC STRAIN=565/56, 84/110-1, AND 83/39;
RA Krejany E.O.;
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
RN
[8]
RP SEQUENCE FROM N.A.
RC STRAIN=RDEC-1;
RX MEDLINE=98254123; PubMed=9593291;
RA Elliott S.J., Wainwright L.A., McDaniel T.K., Jarvis K.G., Deng Y.K.,
RA Lai L.C., McNamara B.P., Donnenberg M.S., Kaper J.B.;
RT "The complete sequence of the locus of enterocyte effacement (LEE)
RT from enteropathogenic Escherichia coli E2348/69.";
RL Mol. Microbiol. 28:1-4(1998).
RN
[9]
RP SEQUENCE FROM N.A.
RC STRAIN=RDEC-1;
RX MEDLINE=21153569; PubMed=11254564;
RA Zhu C., Agin T.S., Elliott S.J., Johnson L.A., Thate T.E., Kaper J.B.,
RA Boedeker E.C.;
RT "Complete Nucleotide Sequence and Analysis of the Locus of Enterocyte
RT Effacement from Rabbit Diarrheagenic Escherichia coli RDEC-1.";
RL Infect. Immun. 69:2107-2115(2001).
DR EMBL; X99670; CAA67984.1; -
DR EMBL; U80796; AAB69980.1; -
DR EMBL; AF059713; AAC64002.1; -
DR EMBL; AF054421; AAC82360.1; -
DR EMBL; AF144010; AAD34584.1; -
DR EMBL; AF144008; AAD34582.1; -
DR EMBL; AF144009; AAD34583.1; -
DR EMBL; AF200363; AAK26729.1; -
SQ SEQUENCE 314 AA; 3320 MW; D9B781454CCA5267 CRC64;

Query Match 13.6%; Score 105.5; DB 2; Length 314;
Best Local Similarity 28.7%; Pred. No. 12;
Matches 45; Conservative 23; Mismatches 68; Indels 21; Gaps 6;

QY 14 GGTTLNLGNAPEKALNRNEERAIDEKQKQA--IEDKEATT---ATEAASDALEALADQ 68
DB 100 GGAISSVLG-----ILGSFAINSATKGSADIAQKTASTSSKADDAASDTATKTLTKA 152
QY 69 T-----DALQSEEAHVVKADNAADALEALADQTDALQSEEAHVVDNAASDAWEKAATP 124
DB 153 TESVADAVEADASSVMOQAMTTATRAASRTSDVADIADSAQSAQSLAENAAADAQAQASRA 212
QY 125 -----IALDVKKTKDTRPVVKKERQNVTLPTTGEES 157
DB 213 SRFMAAVD-KITGSTPPIAVTSIAEGTKTLPTTVSES 248

RESULT 34
Q92IX5 PRELIMINARY; PRT; 389 AA.
AC Q92IX5;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE WP22.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AL168;
RA Thern A., Mustfelt M., Lindahl G.;
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF007557; AAD01409.1; -
DR InterPro; IPR001899; Gram_pos_anchor.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR PRINTS; PR00015; GPOSANCHOR.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
KW Transmembrane.

Query Match 13.6%; Score 105.5; DB 2; Length 389;
Best Local Similarity 27.5%; Pred. No. 16;
Matches 53; Conservative 25; Mismatches 67; Indels 48; Gaps 10;

QY 6 EOPRIIPNGGTLTNLLGNA-----PEKLALNRNEERAIDEK---KKOAIED 47
DB 180 EIAKLQSEAAATLENLLGSAKRELTDLQAKLDATAEAKAKLESQVTTLENLLGSAKRELT 239
QY 48 KEATTATEAASSDALE-----ALADQTDALQSEEA---AVVKADNAASDALE-----ALADQ 96
DB 240 LQAKLDRAAEKEKELQSQAAALEKQLEATKKELADLQAKLVATNQEKLEAEAKALKEQ 299
QY 97 TDALQSEEAHVVDNAASDAWEKAATPIALDVKKTKDTRPVVKKERQNVN----- 148
DB 300 L-AKQAEELAKLKADKASG-----AQKP---DTKPGNKEVTPRPSQRTWTNKASMAQTK 350
QY 149 -TLPTTGEES-SNP 159
||:||||: ||
```





```
AC Q47802; Q52199;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE SURFACE EXCLUSION PROTEIN PRECURSOR (SEAL).
GN SEPI.
OS Enterococcus faecalis (Streptococcus faecalis).
OG Plasmid pPD1.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Enterococcaceae;
OC Enterococcus.
OX NCBI_TaxID=1351;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97074879; PubMed=8917306;
RA Hirt H., Wirth R., Muscholl A.;
RT "Comparative analysis of 18 sex pheromone plasmids from Enterococcus
RT faecalis: detection of a new insertion element on pPD1 and
RT implications for the evolution of this plasmid family.";
RL Mol. Gen. Genet. 252:640-647(1996).
RN [2]
RP SEQUENCE OF 1-127 FROM N.A.
RX MEDLINE=96032393; PubMed=7559344;
RA Nakayama J., Yoshida K., Kobayashi H., Isogai A., Clewell D.B.,
RA Suzuki A.;
RT "Cloning and characterization of a region of Enterococcus faecalis
RT plasmid pPD1 encoding pheromone inhibitor (ipd), pheromone sensitivity
RT (traC), and pheromone shutdown (traB) genes.";
RL J. Bacteriol. 177:5567-5573(1995).
RN [3]
RP SEQUENCE OF 1-127 FROM N.A.
RX MEDLINE=96032394; PubMed=7559345;
RA Fujimoto S., Tomita H., Wakamatsu E., Tanimoto K., Ike Y.;
RT "Physical mapping of the conjugative bacteriocin plasmid pPD1 of
RT Enterococcus faecalis and identification of the determinant related to
RT the pheromone response.";
RL J. Bacteriol. 177:5574-5581(1995).
DR EMBL; X96976; CAA65662.1; -.
DR EMBL; D78016; BAA11202.1; -.
KW Signal; Plasmid.
FT SIGNAL. 1 30 POTENTIAL.
SQ SEQUENCE 843 AA; 93049 MW; 632C8BDC8E79EB2C CRC64;

Query Match 13.4%; Score 104.5; DB 2; Length 843;
Best Local Similarity 28.3%; Pred. No. 41;
Matches 39; Conservative 27; Mismatches 61; Indels 11; Gaps 5;

QY 7 OPRIPNGGTLNLLGNAPEKALRNEERAIDELKKAIEDKEATTAIEAASDALEALA 66
Db 32 QPK-----TPENASSQPTVKATQTTEQAITEKQOOVIE-KQAIVDKQKQVADTAKKE 84

QY 67 DQTD-ALQSEEAHVVKADNAASDALEALADQTDALQSEEAHVQSDNAASDAWEKAATPI 125
Db 85 DTIDQSVKQGVVQDKNDALVQSQAVTQ-QAVVDEAKV--DEATPSAIEKAKEQV 141

QY 126 ALDVKTKTKDPVVKKEE 143
Db 142 ATDTQAVDDQKQVVEQAQ 159

RESULT 39
Q9BHC4 PRELIMINARY; PRT; 924 AA.
AC Q9BHC4;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE POSSIBLE R27-2 PROTEIN (FRAGMENT).
GN P883.45.
OS Leishmania major.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5664;
RN [1]
RP SEQUENCE FROM N.A.
```

```
RC STRAIN=FRIEDLIN;
RA Saunders D., Murphy L., Silva W., Ivens A.C., Quail M.,
RA Rajandream M.A., Barrell B.G.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=FRIEDLIN;
RX MEDLINE=98146435; PubMed=9477341;
RA Ivens A.C., Lewis S.M., Bagherzadeh A., Zhang L., Chan H.M.,
RA Smith D.F.;
RT "A physical map of the Leishmania major Friedlin genome.";
RL Genome Res. 8:135-145(1998).
DR EMBL; AL590734; CAC37243.1; -.
FT NON_TER 924
SQ SEQUENCE 924 AA; 99390 MW; 28F42FB20499E85A CRC64;

Query Match 13.4%; Score 104.5; DB 5; Length 924;
Best Local Similarity 25.9%; Pred. No. 46;
Matches 37; Conservative 24; Mismatches 45; Indels 37; Gaps 4;

QY 5 IEQPIIPNGGTL-----TNLLGNAPEKALRNEERAIDELKKAIEDKEATTAIEAA 57
Db 484 VEEMRSLISGLLTLLHQLVTDKGDVPEAEAPQPSQLERSLSTEEVADQGTSAVDAA 543

QY 58 SSDALEALADQTDALQSEEAHVVKADNAASDALEA-----LADQTDAL 100
Db 544 SSEA-----QKEAALAEAAADEPAETKTKVSLVQGEIDNLADSIADFLKQ 592

QY 101 OSEEA--EVVOSDNAASDAWEKA 121
Db 593 HSQEATLEATASDKADAPAVEEA 615

RESULT 40
Q07290 PRELIMINARY; PRT; 1822 AA.
AC Q07290;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE EF PROTEIN.
GN EFP*.
OS Streptococcus suis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1307;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1890;
RX MEDLINE=93328288; PubMed=8335363;
RA Smith H.E., Reek F.H., Vecht U., Gielkens A.L.J., Smits M.A.;
RT "Repeats in an extracellular protein of weakly pathogenic strains of
RT Streptococcus suis type 2 are absent in pathogenic strains.";
RL Infect. Immun. 61:3318-3326(1993).
DR EMBL; X71880; CAA50714.1; -.
SQ SEQUENCE 1822 AA; 192631 MW; 3838960C77641D7D CRC64;

Query Match 13.4%; Score 104.5; DB 2; Length 1822;
Best Local Similarity 24.5%; Pred. No. 98;
Matches 35; Conservative 32; Mismatches 53; Indels 23; Gaps 5;

QY 16 TLTNLLGNAPEKALRNEERAIDELKKAIEDKEATTAIEAASDALEALADQTD-AL 72
Db 1214 SLDTGTGEARDAVELAKDK----ELAKEATITEEETATKIVEKLAEDTRKAIDENPNLSD 1269

QY 73 OSEEAHVVKADNAASDALEALADQTDALQSEEAHVQSDNAASDAWEKAATPIALDVKTK 132
Db 1270 EDKQAEIKLTDVAKTLATIRDNADK-RTQAEAKQA-----LADLEKA 1313

QY 133 KDKPKVVKKEERQNVNLTPTTGE 155
Db 1314 KETQKIADKAAIDRLTILVKDGE 1336
```

Search completed: October 13, 2002, 04:48:06  
Job time : 74.8717 secs

...

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2002, 03:33:22 ; Search time 22.9572 Seconds  
(without alignments)  
268.169 Million cell updates/sec

Title: US-09-847-539A-6

Perfect score: 777

Sequence: 1 VDSPIEQPRIIPNGGTLTNL.....KKEERQNVNTLPTTGESNP 159

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	188	24.2	448	1 SPGL_STRSP	P06654 streptococ
2	187	24.1	593	1 SPQ2_STRSP	P19909 streptococ
3	131.5	16.9	309	1 SANT_PLAFN	P04928 plasmodium
4	110.5	14.2	774	1 STP_LAMB0	P03764 bacterioph
5	110	14.2	490	1 CE05_ECOLI	Q47500 escherichia
6	110	14.2	490	1 CE10_ECOLI	Q47125 escherichia
7	108.5	14.0	421	1 TOLA_ECOLI	P19934 escherichia
8	107	13.8	1679	1 Y109_YEAST	P40457 saccharomyc
9	105.5	13.6	1120	1 STPR_ECOLI	P76072 escherichia
10	105	13.5	388	1 MRP4_STRPY	P30141 streptococ
11	103	13.3	1189	1 SC11_CHICK	Q90988 gallus gall
12	102	13.1	929	1 E4L3_MOUSE	Q9wv92 mus musculu
13	101.5	13.1	539	1 MY33_HYDAT	P39922 hydra atten
14	101.5	13.1	1296	1 ASAL_ENTFA	P17953 enterococcu
15	101	13.0	771	1 CALD_CHICK	P12957 gallus gall
16	100	12.9	857	1 NFM_CHICK	P16053 gallus gall
17	100	12.9	1637	1 MRSP_STAAU	P80544 staphylococ
18	99.5	12.8	1337	1 DEXT_STRDO	P39653 streptococ
19	99	12.7	464	1 AK15_RAT	P24587 ratus norv
20	99	12.7	896	1 IF2_KLEFX	Q9zf28 klebsiella
21	98.5	12.7	407	1 IE68_HSVSA	Q01042 herpesvirus
22	98.5	12.7	897	1 IF2_ENTCL	Q9zf25 enterobacte
23	98.5	12.7	1290	1 XCPX_XENLA	P50532 xenopus lae
24	98	12.6	241	1 NEUM_SERCA	Q98987 serinus can
25	96	12.4	1714	1 SYEP_DROME	P28668 drosophila
26	95.5	12.3	708	1 ICAL_HUMAN	P20810 homo sapien
27	95.5	12.3	740	1 RNFC_ECOLI	P77611 escherichia
28	95	12.2	198	1 HBHA_MYCTU	Q11142 mycobacteri
29	95	12.2	372	1 TOLA_HAEIN	P44678 haemophilus
30	95	12.2	483	1 M6_STRPY	P08089 streptococ
31	95	12.2	564	1 M12_STRPY	P19401 streptococ
32	94.5	12.2	1077	1 HLES_DROME	Q02308 drosophila
33	93.5	12.0	166	1 VP06_BPPRD	P27379 bacterioph

#### RESULT 1

ID	SPGL_STRSP	STANDARD;	PRT;	448 AA.
AC	P06654:			
DT	01-JAN-1988 (Rel. 06, Created)			
DT	01-JAN-1988 (Rel. 06, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Immunoglobulin G binding protein G precursor (IGG binding protein G).			
GN	SPG.			
OS	Streptococcus sp. (Lancefield group G).			
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;			
OC	Streptococcus.			
OX	NCBI_TaxID=1306;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=86304178; PubMed=3745123;			
RA	Pahnestock S.R., Alexander P., Nagle J., Filpula D.;			
RT	"Gene for an immunoglobulin-binding protein from a group G streptococcus.";			
RT	streptococcus.";			
RL	J. Bacteriol. 167:870-880(1986).			
RN	[2]			
RP	X-RAY CRYSTALLOGRAPHY (2.07 ANGSTROMS) OF 228-282.			
RX	MEDLINE=94213848; PubMed=8161530;			
RA	Gallagher T., Alexander P., Bryan P., Gilliland G.L.;			
RT	"Two crystal structures of the B1 immunoglobulin-binding domain of streptococcal protein G and comparison with NMR.";			
RL	Biochemistry 33:4721-4729(1994).			
RN	[3]			
RP	X-RAY CRYSTALLOGRAPHY (1.1 ANGSTROMS) OF 293-351.			
RX	MEDLINE=95055731; PubMed=7966308;			
RA	Derrick J.P., Wigley D.B.;			
RT	"The third IgG-binding domain from streptococcal protein G. An analysis by X-ray crystallography of the structure alone and in a complex with Fab.";			
RL	J. Mol. Biol. 243:906-918(1994).			
RN	[4]			
RP	X-RAY CRYSTALLOGRAPHY (3.5 ANGSTROMS) OF 297-352.			
RX	MEDLINE=95308043; PubMed=7788293;			
RA	Sauer-Eriksson A.E., Kleywegt G.J., Uhlen M., Jones T.A.;			
RT	"Crystal structure of the C2 fragment of streptococcal protein G in complex with the Fc domain of human IgG.";			
RL	Structure 3:265-278(1995).			
RN	[5]			
RP	X-RAY CRYSTALLOGRAPHY (1.1 ANGSTROMS) OF 292-352.			
RA	Butterworth S., Lamzin V.S., Wigley D.B., Derrick J.P., Wilson K.S.;			
RT	Submitted (APR-1997) to the PDB data bank.			
RL				
RN	[6]			
RP	STRUCTURE BY NMR OF 298-351.			
RX	MEDLINE=91335209; PubMed=1871600;			
RA	Gronenborn A.M., Filpula D.R., Essig N.Z., Achari A., Whitlow M.,			
RT	Wingfield P.T., Clore G.M.;			
RL	"A novel, highly stable fold of the immunoglobulin binding domain of streptococcal protein G.";			
RT	streptococcal protein G.";			
RL	Science 253:657-661(1991).			
CC	-!- FUNCTION: BINDS TO THE CONSTANT FC REGION OF IGG WITH HIGH AFFINITY.			
CC				

P13692 enterococcu  
Q08641 saccharomyc  
P26041 mus musculu  
O74191 schizosacch  
Q14203 homo sapien  
O9hp81 halobacteri  
P71410 halobacteri  
P15205 ratus norv  
P15335 homo sapien  
P12759 chlamydomon  
P14314 homo sapien  
Q9hr92 halobacteri

#### ALIGNMENTS

```

CC -!- SUBCELLULAR LOCATION: Type I membrane protein. Cell wall.
CC -!- SIMILARITY: TO OTHER STREPTOCOCCAL AND STAPHYLOCOCCAL PROTEINS
CC IN THE REGION OF THE MEMBRANE ANCHOR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M13825; AAA03664.1; -.
DR PIR; A24496; A24496.
DR PDB; 1PGA; 30-APR-94.
DR PDB; 1PGB; 30-APR-94.
DR PDB; 1PGX; 15-JUL-92.
DR PDB; 1IGC; 03-JUN-95.
DR PDB; 1IGD; 01-NOV-94.
DR PDB; 2IGD; 29-JUL-98.
DR PDB; 1PCC; 20-JUL-95.
DR PDB; 1GB1; 15-APR-93.
DR InterPro; IPR002988; GA.
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR000724; Igg_bind_B.
DR Pfam; PF01468; GA; 2.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR Pfam; PF01378; Igg_binding_B; 2.
DR PRINTS; PR00015; GP05ANCHOR.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; 1.
KW Igg-binding protein; Repeat; Transmembrane; Cell wall; Signal;
KW 3D-structure.
FT SIGNAL 1 33
FT CHAIN 34 448 IMMUNOGLOBULIN G BINDING PROTEIN G.
FT DOMAIN 34 422 EXTRACELLULAR.
FT DOMAIN 423 443
FT DOMAIN 444 448
FT DOMAIN 69 111 CYTOPLASMIC.
FT DOMAIN 104 215 ALA-RICH.
FT REPEAT 104 140 2 X 37 AA REPEATS.
FT REPEAT 179 215 1-1.
FT DOMAIN 228 352 1-2.
FT REPEAT 228 282 2 X 55 AA REPEATS.
FT REPEAT 298 352 2-1.
FT DOMAIN 386 410 5 X 5 AA REPEATS OF [DE]-D-A-K-K.
FT DOMAIN 414 419 CONSERVED IN GRAM-POSITIVE COCCI SURFACE
FT PROTEINS.
SQ SEQUENCE 448 AA; 47567 MW; A0759060C8F9E6CA CRC64;

Query Match 24.2%; Score 188; DB 1; Length 448;
Best Local Similarity 38.9%; Pred. No. 9.4e-06;
Matches 61; Conservative 20; Mismatches 60; Indels 16; Gaps 6;

Qy 1 VDSPIEQRIPTNGTLNLGNPEKIALNEERAIDELKQKATIEDKEATTAIEAASD 60
   ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 34 VDSPIEDPIIRNGGELTNLLGNSETTLALRNEESATADLTAAAVADVVAAMAAENAGAA 93

Qy 61 ALEALADQTDALQSEEAHVAKDN--AASDALEAL--ADQTDALQSEEAHVQSDNAA- 114
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 94 AWEAAA--AADALAKAKADALKEFNKYGSDYKKNLNNAKTVEGIKDLQAQVVEAKKAR 152

Qy 115 -SDAWE-----KAATPIALQVTKTK--DTKPVVKKE 142
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 153 ISEATDGLSLDFKLSQTPAEDTVKSLAEAKVLANRE 189

RESULT 2
SPG2_STRSP
ID SPG2_STRSP STANDARD: PRT; 593 AA.
AC SP1909;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)

```

```

DE GN Streptococcus sp. (strain G148).
OS Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1306;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=G148;
RX MEDLINE=88029445; PubMed=3665928;
RA Olsson A., Eliasson M., Guss B., Nilsson B., Hellman U.,
RT Lindberg M., Uhlen M.;
RT "Structure and evolution of the repetitive gene encoding
RT streptococcal protein G.";
RL Eur. J. Biochem. 168:319-324(1987).
[2]
RN SEQUENCE FROM N.A.
RC STRAIN=GX7805;
RX MEDLINE=88015586; PubMed=3658689;
RA Filipula D., Alexander P., Fahnestock S.R.;
RT "Nucleotide sequence of the protein G gene from Streptococcus GX7805,
RT and comparison to previously reported sequences.";
RL Nucleic Acids Res. 15:7210-7210(1987).
[3]
RN SEQUENCE OF 114-593 FROM N.A.
RC STRAIN=G148;
RX MEDLINE=96300657; PubMed=3017704;
RA Guss B., Eliasson M., Olsson A., Uhlen M., Frej A.-K., Joernvall H.,
RT Flock J.-I., Lindberg M.;
RT "Structure of the Igg-binding regions of streptococcal protein G.";
RL EMBO J. 5:1567-1575(1986).
[4]
RN STRUCTURE BY NMR OF 371-427.
RC STRAIN=G148;
RX MEDLINE=98290449; PubMed=9628485;
RA Malakauskas S.M., Mayo S.L.;
RT "Design, structure and stability of a hyperthermophilic protein
RT variant.";
RL Nat. Struct. Biol. 5:470-475(1998).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein. Cell wall.
CC -!- SIMILARITY: TO OTHER STREPTOCOCCAL AND STAPHYLOCOCCAL PROTEINS
CC IN THE REGION OF THE MEMBRANE ANCHOR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X06173; CAA29540.1; -.
DR EMBL; Y00428; CAA68489.1; -.
DR EMBL; X04015; CAA27638.1; -.
DR PIR; S00128; S00128.
DR PIR; A26314; A26314.
DR PDB; 1GB4; 22-JUL-98.
DR InterPro; IPR002988; GA.
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR000724; Igg_bind_B.
DR Pfam; PF01468; GA; 3.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR Pfam; PF01378; Igg_binding_B; 3.
DR PRINTS; PR00015; GP05ANCHOR.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; 1.
KW Igg-binding protein; Repeat; Transmembrane; Cell wall; Signal;
KW 3D-structure.
FT SIGNAL 1 33
FT CHAIN 34 593 IMMUNOGLOBULIN G BINDING PROTEIN G.
FT DOMAIN 34 567 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 568 588 MEMBRANE ANCHOR.
FT DOMAIN 589 593 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 69 111 ALA-RICH.

```





Db 208 KKLAEAKELAKABSDVQSKAIVSRVAGELENAQK-----SYDVKVTGPGWRDYOQK 261  
 QY 142 EERQ 145  
 Db 262 LERQ 265  
 III  
 RESULT 7  
 TOLA\_ECOLI STANDARD; PRT; 421 AA.  
 AC P19934; 1991 (Rel. 17, Created)  
 DT 01-FEB-1991 (Rel. 17, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Tola protein.  
 GN TOLA OR CIM OR EXCC OR LXY OR B0739.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-JM105;  
 RX MEDLINE=90078104; PubMed=2687247;  
 RA Levensgood S.K., Webster R.E.;  
 RT "Nucleotide sequences of the tola and tolB genes and localization of  
 their products, components of a multistep translocation system in  
 Escherichia coli.";  
 RL J. Bacteriol. 171:6600-6609(1989).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K12 / MG1655;  
 RX MEDLINE=97426617; PubMed=9278503;  
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RT "The complete genome sequence of Escherichia coli K-12.";  
 RL Science 277:1453-1474(1997).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K12;  
 RX MEDLINE=97061202; PubMed=8905232;  
 RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,  
 RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,  
 RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,  
 RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,  
 RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,  
 RA Yano M., Horiuchi T.;  
 RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome  
 corresponding to the 12.7-28.0 min region on the linkage map.";  
 RL DNA Res. 3:137-155(1996).  
 RN [4]  
 RP DOMAINS.  
 RX MEDLINE=91296736; PubMed=2068069;  
 RA Levensgood S.K., Beyer W.F. Jr., Webster R.E.;  
 RT "Tola; a membrane protein involved in colicin uptake contains an  
 extended helical region.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 88:5939-5943(1991).  
 RN [5]  
 RP INTERACTION WITH PORINS.  
 RX MEDLINE=97133271; PubMed=8978668;  
 RA Derouiche R., Gavioli M., Benedetti H., Prillipov A., Lazdunski C.,  
 RA Lioubes R.;  
 RT "Tola central domain interacts with Escherichia coli porins.";  
 RL EMBO J. 15:6408-6415(1996).  
 RN [6]  
 RP X-RAY CRYSTALLOGRAPHY (1.85 ANGSTROMS) OF 298-421.  
 RX MEDLINE=99332679; PubMed=10404600;  
 RA Lubkowski J., Hennecke F., Plueckthun A., Wlodawer A.;  
 RT "Filamentous phage infection: crystal structure of g3p in complex  
 with its coreceptor, the C-terminal domain of Tola.";  
 RT Structure 7:711-722(1999).

CC -!- FUNCTION: INVOLVED IN THE TONB-INDEPENDENT UPTAKE OF GROUP A  
 CC COLICINS (COLICINS A, E1, E2, E3, AND K). NECESSARY FOR THE  
 CC COLICINS TO REACH THEIR RESPECTIVE TARGETS AFTER INITIAL  
 CC BINDING TO THE BACTERIA. ALSO INVOLVED IN THE TRANSLOCATION  
 CC OF BACTERIOPHAGE DNA.  
 CC -!- SUBUNIT: INTERACTS, VIA DOMAIN II, WITH PORINS OMPC, OMPC, PHOE  
 CC AND LAMB.  
 CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Inner membrane.  
 CC  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; M28232; AAA24683.1; -;  
 CC EMBL; AE000177; AAC73833.1; -;  
 CC EMBL; D90713; BAA35405.1; -;  
 CC PIR; JY0057; JY0057;  
 CC PDB; 1TOL; 20-MAY-99.  
 CC Ecogene; EG11007; tola.  
 CC Transport; Protein transport; Bacteriocin transport; Transmembrane;  
 CC Repeat; Inner membrane; 3D-structure; Complete proteome.  
 CC DOMAIN 1 13 CYTOPLASMIC (POTENTIAL).  
 CC TRANSMEM 14 34 POTENTIAL.  
 CC DOMAIN 35 421 PERIPLASMIC (POTENTIAL).  
 CC DOMAIN 48 310 DOMAIN II (ALPHA-HELICAL).  
 CC DOMAIN 311 421 DOMAIN III (FUNCTIONAL).  
 CC DOMAIN 224 278 10 X TANDEM REPEATS OF [ED]-K(1,2)-  
 CC A(2,4).  
 CC SQ SEQUENCE 421 AA; 43156 MW; 8B2F52B4B97C655E CRC64;  
 Query Match 14.0%; Score 108.5; DB 1; Length 421;  
 Best Local Similarity 32.6%; Pred. No. 1.4;  
 Matches 42; Conservative 25; Mismatches 43; Indels 19; Gaps 6;  
 QY 26 EKALRNEERAIDELKKQA-IEDKEATTAIEAASDDALEALADQTDALQSEEAIVVKN 84  
 Db 109 ERLAAEQKKQAEAEAKQAEKQAEFAAKAAADA-KAKA-EADAKAEAE-----AKK 162  
 QY 85 AASDA---LEALADQTDALQSEEAIVV-----OSDNAASDAWEKAATPIALDVKK 132  
 Db 163 AAADAKKAEAEAKAEAKAEAKAEAKAEAKAEAKAEAKAEAKAEAKAEAKAEAKAE 222  
 QY 133 KDTKPVVKK 141  
 Db 223 AEKKAEEK 231  
 RESULT 8  
 Y109\_YEAST STANDARD; PRT; 1679 AA.  
 AC P40457;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 01-FEB-1995 (Rel. 31, Last annotation update)  
 DE Hypothetical 195.1 kDa protein in DNA43-UBII intergenic region.  
 GN Y11149C.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetaceae; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-S288C / AB972;  
 RA Barrell B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,  
 RA Churcher C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,  
 RA Gentles S., Hamlyn N., Horsnell T.S., Hunt S., Jagels K., Jones M.,  
 RA Lewis E., Lyne G., Moule S., Moule T., Odell C., Pearson D.,  
 RA Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V.,  
 RA Walsh S.V., Whitehead S.;

```
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; 238059; CA86129.1; .
DR PIR; S48385; S48385.
DR SGD; S000141; MLP2.
DR KW Hypothetical protein.
SQ SEQUENCE 1679 AA; 195141 MW; 298950CC52202D8F CRC64;
-----
Query Match 13.8%; Score 107; DB 1; Length 1679;
Best Local Similarity 26.8%; Pred. No. 7.7;
Matches 44; Conservative 32; Mismatches 52; Indels 36; Gaps 7;
QY 17 LTNLLGNAPEKL-----ALRNEERAIDELKQAIEDKEATTAEIASSDALEALADQTD 71
DB 1261 LTNELSDLGKGLSSRENADLENFNRLLKQAEKLDASKKQQAALTNELNELKAIRK 1320
QY 72 LQS-----ERAAVVKAD-----NAASDALEALADQTDALQSEAEVQSDN 112
DB 1321 LEQDLHFENAKVLDITLKLKAHELQSEDSYRDHEKDYRTLMETESLK-KELQIFKTAN 1379
QY 113 AASDAWEKAATPIALDVKTKD-----TKPVVKK-EERQNVNT 149
DB 1380 SSSDAPEK-----LKNVMEKEKRIIDERTKEFEKLIQETLNKST 1419
-----
RESULT 9
STPR_ECOLI
ID STPR_ECOLI STANDARD; PRT; 1120 AA.
AC P76072; P77560;
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Slide tail fiber protein homolog from lambdaoid prophage Rac.
GN STFR OR B1372.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RA "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE=97251357; PubMed=9097039;
RA Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
RA Makino T., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T., Saico N.,
RA Sempel G., Seki Y., Sivasubram S., Tagami H., Takeda J.,
RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horiuchi T.;
RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 28.0-40.1 min region on the linkage map.";
RL DNA Res. 3:363-377(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE=92409576; PubMed=1528877;
RA O'Toole P., Stenberg L., Rissler M., Lindahl G.;
RT "Two major classes in the M protein family in group A streptococci.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:8661-8665(1992).
CC -|- FUNCTION: BINDS IGG MOLECULES OF THE IGI, IG2 AND IG4 SUBCLASSES,
CC AND ALSO BINDS FIBRINOGEN.
CC -|- SUBCELLULAR LOCATION: Type I membrane protein. Cell wall.
CC -|- SIMILARITY: TO OTHER STREPTOCOCCAL AND STAPHYLOCOCCAL PROTEINS
CC IN THE REGION OF THE MEMBRANE ANCHOR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
```

```
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF000234; AAC74454.1; ALT_INIT.
DR EMBL; D90774; BAA14966.1; .
DR EMBL; D90775; BAA14975.1; .
DR HSSP; P04002; IWFA.
DR ECoGene; EGI3370; stfr.
DR InterPro; IPR004089; Chemotaxis_transducer.
DR KW Hypothetical protein; Fiber protein; Repeat; Complete proteome.
SQ SEQUENCE 1120 AA; 113779 MW; 542E59D71EE795B4 CRC64;
-----
Query Match 13.6%; Score 105.5; DB 1; Length 1120;
Best Local Similarity 31.5%; Pred. No. 6.3;
Matches 41; Conservative 18; Mismatches 58; Indels 13; Gaps 4;
QY 15 GTLTNLLGNAPE-----KLALRNEERAIDELKQAIEDKEATTAEIASSD-----ALEALA 66
DB 84 GTLNDFLGAMTEDDARPEALRRFELMWVEHVARNASAVAQNTAAAKKSADASTSAREANT 143
QY 67 DQTDALQSEAAVVKADNAASDALEALADQ-TDALQSEAEVQSDNAAASDAWEKAATPI 125
DB 144 HAADAADSARAASTSAGQAASSAQSSAGTASTKATEA-----SKSAAAASSKSAAT 199
QY 126 ALDVKTKD 135
DB 200 SAGAAKTSET 209
-----
RESULT 10
MRP4_STRPY
ID MRP4_STRPY STANDARD; PRT; 388 AA.
AC P30141;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DE Fibrinogen- and Ig-binding protein precursor (MRP protein).
GN MRP4.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-AP4 / SEROTYPE M4;
RX MEDLINE=92269579; PubMed=1588817;
RA Stenberg L., O'Toole P., Lindahl G.;
RT "Many group A streptococcal strains express two different
RT immunoglobulin-binding proteins, encoded by closely linked genes.
RT characterization of the proteins expressed by four strains of
RT different M-type.";
RL Mol. Microbiol. 6:1185-1194(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-AP4 / SEROTYPE M4;
RX MEDLINE=92409576; PubMed=1528877;
RA O'Toole P., Stenberg L., Rissler M., Lindahl G.;
RT "Two major classes in the M protein family in group A streptococci.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:8661-8665(1992).
CC -|- FUNCTION: BINDS IGG MOLECULES OF THE IGI, IG2 AND IG4 SUBCLASSES,
CC AND ALSO BINDS FIBRINOGEN.
CC -|- SUBCELLULAR LOCATION: Type I membrane protein. Cell wall.
CC -|- SIMILARITY: TO OTHER STREPTOCOCCAL AND STAPHYLOCOCCAL PROTEINS
CC IN THE REGION OF THE MEMBRANE ANCHOR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
```























RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirska R., Tector C., Turner K., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.:  
RT "The genome sequence of *Drosophila melanogaster*."  
RL Science 287:2185-2195(2000).  
CC -|- CATALYTIC ACTIVITY: ATP + L-glutamate + tRNA(Glu) = AMP +  
CC -|- CATALYTIC ACTIVITY: ATP + L-proline + tRNA(Pro) = AMP +  
CC -|- CATALYTIC ACTIVITY: L-prolyl-tRNA(Pro).  
CC -|- SUBUNIT: COMPONENT OF THE MULTISYNTHETASE COMPLEX WHICH IS  
CC COMPRISED OF A BIFUNCTIONAL GLUTAMYL-PROLYL-TRNA SYNTHETASE, THE  
CC MONOSPECIFIC ISOLEUCYL-, LEUCYL-, GLUTAMINYL-, METHIONYL-, LYSYL-,  
CC ARGINYL-, AND ASPARTYL-TRNA SYNTHETASES AS WELL AS THREE AUXILIARY  
CC PROTEINS, P18, P48 AND P43.  
CC -|- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO CLASS-II  
CC AMINOACYL-TRNA SYNTHETASE FAMILY.  
CC -|- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO CLASS-II  
CC AMINOACYL-TRNA SYNTHETASE FAMILY.  
CC -|- SIMILARITY: CONTAINS 6 "WHEP-TRS" DOMAINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M74104; AAA28594.1; -;  
DR EMBL; U59923; AAC47469.1; -;  
DR EMBL; AE003745; AAF56211.1; -;  
DR PIR; S18644; S18644.  
DR HSP; P00962; IGTR.  
DR FlyBase; FBgn0005674; Aats-glupro.  
DR InterPro; IPR002106; AA\_trna\_ligase\_II.  
DR InterPro; IPR004046; GST\_C.  
DR InterPro; IPR004154; HGTP\_anticonodon.  
DR InterPro; IPR000738; WHEP-TRS.  
DR InterPro; IPR000924; trna-synt\_lc.  
DR InterPro; IPR002314; trna-synt\_2b.  
DR InterPro; IPR001412; trna-synt\_I.  
DR InterPro; IPR002316; trna-synt\_pro.  
DR Pfam; PF00043; GST-C; 1.  
DR Pfam; PF03129; HGTP\_anticonodon; 1.  
DR Pfam; PF00749; trna-synt\_lc; 1.  
DR Pfam; PF00587; trna-synt\_2b; 1.  
DR Pfam; PF00458; WHEP-TRS; 6.  
DR Pfam; PF00458; WHEP-TRS; 6.  
DR PRINTS; PR00987; TRNASYNTHGLU.  
DR PRINTS; PR01046; TRNASYNTHGLU.  
DR PROSITE; PS00178; AA\_TRNA\_LIGASE\_I; 1.  
DR PROSITE; PS00179; AA\_TRNA\_LIGASE\_II\_1; FALSE\_NEG.  
DR PROSITE; PS00339; AA\_TRNA\_LIGASE\_II\_2; FALSE\_NEG.  
DR PROSITE; PS00762; WHEP-TRS; 6.  
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;  
KW Multifunctional enzyme; Repeat.  
FT DOMAIN 170 754 GLUTAMYL-TRNA SYNTHETASE.  
FT DOMAIN 755 800 WHEP-TRS 1.  
FT DOMAIN 827 872 WHEP-TRS 2.  
FT DOMAIN 901 946 WHEP-TRS 3.  
FT DOMAIN 980 1025 WHEP-TRS 4.  
FT DOMAIN 1055 1100 WHEP-TRS 5.  
FT DOMAIN 1129 1173 WHEP-TRS 6.  
FT DOMAIN 1174 1180 POLY-GLY.  
FT DOMAIN 1207 1714 POLY-GLY-TRNA SYNTHETASE.  
FT SITE 209 220 "HIGH" REGION.  
FT SITE 438 442 "KMSKS" REGION.

FT BINDING 441 ATP (BY SIMILARITY).  
FT CONFLICT 102 TSPLP -> DKSIA (IN REF. 3).  
FT CONFLICT 233 VC -> AF (IN REF. 3).  
FT CONFLICT 341 NTACA -> KYCVR (IN REF. 3).  
FT CONFLICT 583 K -> R (IN REF. 3).  
FT CONFLICT 592 L -> A (IN REF. 3).  
FT CONFLICT 692 T -> S (IN REF. 3).  
FT CONFLICT 753 T -> S (IN REF. 3).  
FT CONFLICT 802 T -> T (IN REF. 3).  
FT CONFLICT 873 P -> T (IN REF. 3).  
FT CONFLICT 887 G -> V (IN REF. 3).  
FT CONFLICT 1201 P -> PA (IN REF. 3).  
FT CONFLICT 1461 MISSING (IN REF. 3).  
FT CONFLICT 1587 G -> V (IN REF. 3).  
SQ SEQUENCE 1714 AA; 189197 MW; 6FE8C58045E48ABC CRC64;  
Query Match 12.4%; Score 96; DB 1; Length 1714;  
Best Local Similarity 25.2%; Pred. No. 42;  
Matches 39; Conservative 27; Mismatches 57; Indels 32; Gaps 6;  
QY 28 LALRNEERAI-----DELKKOATEDKEATTAIEAASSDALEALADOTDALQSEEAAYV 80  
DB 1080 LALKGEYKTLGSKDWTTPDAKSEPAVVKKEASPV--SMASPAKDELTOEINA--OGEKYRAA 1136  
QY 81 KADNAASDALEALADQTDALQSEEAQVQSD-----NAASDAWEKAATP 124  
DB 1137 KGNKAAKEVIDAEVAKLLALKAKYKEVTGTDFPVAGRGGGGGGSAKKAPKAPKAPK 1196  
QY 125 IALDVKKTKDTPKPVKKERQNVNLTPTTGESNP 159  
DB 1197 VR---KEPADASGAVKKOTRLG---LEATKEDNLP 1225  
RESULT 26  
ICAL\_HUMAN  
ID ICAL\_HUMAN STANDARD; PRT; 708 AA.  
AC P20810;  
DT 01-FEB-1991 (Rel. 17, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Calpain inhibitor (Calpastatin) (Sperm BS-17 component).  
GN CAST.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91124109; PubMed=2577276;  
RA Asada K., Ishino Y., Shimada M., Shimojo T., Endo M., Kimizuka F.,  
RA Kato I., Maki M., Hatanaka M., Murachi T.;  
RT "cDNA cloning of human calpastatin: sequence homology among human,  
RT pig, and rabbit calpastatins.";  
RL J. Enzym. Inhib. 3:49-56(1989).  
RN [2]  
RP SEQUENCE OF 1-283 FROM N.A., AND ALTERNATIVE SPLICING.  
RC TISSUE=Lung;  
RX MEDLINE=92235069; PubMed=1569094;  
RA Lee W.J., Ma H., Takano E., Yang H.Q., Hatanaka M., Maki M.;  
RT "Molecular diversity in amino-terminal domains of human calpastatin  
RT by exon skipping.";  
RL J. Biol. Chem. 267:8437-8442(1992).  
RN [3]  
RP SEQUENCE OF 101-317 FROM N.A.  
RX MEDLINE=90037002; PubMed=2553724;  
RA Maki M., Bagci H., Hamaguchi K., Ueda M., Murachi T., Hatanaka M.;  
RT "Inhibition of calpain by a synthetic oligopeptide corresponding to  
RT an exon of the human calpastatin gene.";  
RN J. Biol. Chem. 264:18866-18869(1989).  
RN [4]  
RP SEQUENCE OF 125-283 FROM N.A.  
RX MEDLINE=90165962; PubMed=2407243;  
RA Uemori T., Shimojo T., Asada K., Asano T., Kimizuka F., Kato I.,  
RA Maki M., Hatanaka M., Murachi T., Hanzawa H., Arata Y.;

RT \*Characterization of a functional domain of human calpastatin.\*;  
 RL Biochem. Biophys. Res. Commun. 166:1485-1493(1990).  
 RN [5]  
 RP SEQUENCE OF 153-708 FROM N.A. (SHORT ISOFORM), AND REVISIONS.  
 RA Wang L.F.; Wei S.G., Miao S.F., Liu Q.Y., Koide S.S.;  
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.  
 RN [6]  
 RP SEQUENCE OF 523-708 FROM N.A.  
 RN [7]  
 RP SEQUENCE OF 425-708 FROM N.A.  
 RN [8]  
 RP SEQUENCE OF 244-708 FROM N.A.  
 RP TISSUE-Placenta;  
 RC El-Anine M., Talbot G., Despres N., Asselin C., Boire G.,  
 RA Menard H.A.;  
 RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.  
 RN [9]  
 RP PHOSPHORYLATION.  
 RX MEDLINE-91139699; PubMed-1995645;  
 RA Adachi Y., Ishida-Takahashi A., Takahashi C., Takano E., Murachi T.,  
 RA Hatanaka M.;  
 RT "Phosphorylation and subcellular distribution of calpastatin in human  
 RT hemocytoblastic system cells.\*;  
 RL J. Biol. Chem. 266:3968-3972(1991).  
 CC -1- FUNCTION: SPECIFIC INHIBITION OF CALPAIN (CALCIUM-DEPENDENT  
 CC CYSTEINE PROTEASE). PLAYS A KEY ROLE IN POSTMORTEM TENDERIZATION  
 CC OF MEAT AND HAVE BEEN HYPOTHESIZED TO BE INVOLVED IN MUSCLE  
 CC PROTEIN DEGRADATION IN LIVING TISSUE.  
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A  
 CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.  
 CC -1- DOMAIN: HAS FOUR INHIBITORY DOMAINS.  
 CC -1- PTM: THE N-TERMINUS IS BLOCKED.  
 CC -1- PTM: PHOSPHORYLATED.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; D16217; BAA03747.1; -;  
 DR EMBL; D50827; BAA09438.1; -;  
 DR EMBL; M86258; AAB59398.1; -;  
 DR EMBL; M28230; AAA52066.1; -;  
 DR EMBL; M28227; AAA52066.1; JOINED.  
 DR EMBL; M28228; AAA52066.1; JOINED.  
 DR EMBL; M28229; AAA52066.1; JOINED.  
 DR EMBL; M33328; AAA52296.1; -;  
 DR EMBL; U76724; AAC50136.2; -;  
 DR EMBL; S73329; AAB32311.1; -;  
 DR EMBL; U38525; AAA80684.1; -;  
 DR EMBL; U31345; AAB60371.1; -;  
 DR EMBL; U31346; AAB60372.1; -;  
 DR PIR; A34428; A34428.  
 DR MIM; 114090; -;  
 DR InterPro: IPR001259; Calpain\_inhib.  
 DR Pfam: PF00748; Calpain\_inhib; 4.  
 KW Repeat; Thiol protease inhibitor; Alternative splicing;  
 KW Phosphorylation.

FT DOMAIN 1 169 L.  
 FT REPEAT 170 222 INHIBITORY DOMAIN 1.  
 FT REPEAT 304 356 INHIBITORY DOMAIN 2.  
 FT REPEAT 446 499 INHIBITORY DOMAIN 3.  
 FT REPEAT 583 636 INHIBITORY DOMAIN 4.  
 FT VARSPLIC 212 224 MISSING (IN SHORT ISOFORM).  
 FT VARIANT 592 592 G -> E.  
 FT /FTID=VAR\_005298.  
 FT CONFLICT 467 467 R -> L (IN REF. 5).  
 FT CONFLICT 486 488 VKD -> GRE (IN REF. 5).  
 FT CONFLICT 543 543 V -> L (IN REF. 5).  
 FT CONFLICT 562 562 MISSING (IN REF. 8).  
 SQ SEQUENCE 708 AA; 76484 MW; CACD759C9284E3EA CRC64;  
 Query Match 12.3%; Score 95.5; DB 1; Length 708;  
 Best Local Similarity 27.3%; Pred. No. 17; Mismatches 35; Gaps 7;  
 Matches 45; Conservative 21;  
 QY 20 LLGNAPEKALRNEERAIDEL-----KQAIEDKEATTAIEAASSDALEALADQTD 70  
 DB 352 LLPEPEKPKRSESELIDELSEDFDRSECKEKPSTKTEESKAAAPVSEAVSRTS 411  
 QY 71 --ALQS--EAAVVKADNAASDALEALADQTDALQSEAEVQSDNAASDAWEKA----- 121  
 DB 412 MCSIQSAPPEPATLKG-TVPDDAVEALA---DSLGRKEADPEDGKPYMDKVKKAKEEDR 467  
 QY 122 -----ATPIALDVKTKTKPKVVKKEERQNVNLTPTTCEE 156  
 DB 468 EKLGEKEETIPDYRLLEVVDKDGKPLPKESKEQ----LPPMSSED 509  
 RESULT 27  
 RNFC\_ECOLI STANDARD; PRT; 740 AA.  
 ID RNFC\_ECOLI  
 AC P77611;  
 DT 01-MAR-2002 (Rel. 41, Created)  
 DT 01-MAR-2002 (Rel. 41, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Electron transport complex protein rnfc.  
 GN RNFC OR B1629.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OC NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MGI655;  
 RX MEDLINE-97426617; PubMed-9278503;  
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RT "The complete genome sequence of Escherichia coli K-12.\*;  
 RL Science 277:1453-1474(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE-97251357; PubMed-9097039;  
 RA Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,  
 RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,  
 RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,  
 RA Nakade S., Nakamura Y., Nashimoto H., Nishio Y., Oshima T., Saito N.,  
 RA Sampei G., Seki Y., Sivasubram S., Tagami H., Takeda J.,  
 RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horiuchi T.;  
 RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome  
 RT corresponding to the 28.0-40.1 min region on the linkage map.\*;  
 RL DNA Res. 3:363-377(1996).  
 CC -1- FUNCTION: May be part of a membrane complex involved in electron  
 CC transport (By similarity).  
 CC -1- COFACTOR: Binds 2 4FE-4S clusters (Potential).  
 CC -1- SUBUNIT: Composed of at least six subunits; rnfa, rnfb, rnfc,  
 CC rnfd, rnfe and rnfg (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Inner membrane-associated (By similarity).

```
CC CC -!- SIMILARITY: BELONGS TO THE 4FE4S BACTERIAL-TYPE FERREDOXIN FAMILY.
CC RNFC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AE000258; AAC74701.1; -.
CC EMBL: D90806; BAA15384.1; -.
CC EMBL: D90807; BAA15391.1; -.
CC EMBL: D90808; BAA15414.1; -.
CC HSP: P00198; 1FCA.
CC EcoGene: EG13935; rnfC.
CC InterPro: IPR001450; 4Fe4S-ferredoxin.
CC Pfam: PF01512; Complex1_51K; 1.
CC Pfam: PF00037; fer4; 2.
CC PROSITE: PS00198; 4FE4S_FERREDOXIN; 2.
KW Electron transport; Iron-sulfur; 4Fe-4S; Inner membrane;
KW Complete proteome.
FT METAL 377 377 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 380 380 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 383 383 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 387 387 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 416 416 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 419 419 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 422 422 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 426 426 IRON-SULFUR (4FE-4S) (POTENTIAL).
SQ SEQUENCE 740 AA; 80171 MW; 381506475CD6E01F CRC64;

Query Match 12.3%; Score 95.5; DB 1; Length 740;
Best Local Similarity 26.5%; Pred. No. 18;
Matches 40; Conservative 29; Mismatches 45; Indels 37; Gaps 7;

QY 26 EKALNRERATDELK-----KOATEDKEATTAIEAASSDALEALADQTDAL----- 72
Db 442 ETAAIQEKRAAEAKARFEARQARLERKAARLKHSAAVOPAKOKDAATAALARYK 501
QY 73 ----OSEEAAVYKA-----DN-----AASDALEALADQTDALQSEAEVVSQDAAADAWEK 120
Db 502 EKQAOATQPIVAKGERPDNSAIIAAREAKQA-----RAKQAELOQTNDAAATVA-DP 554
QY 121 AATPALDVKKTKDKTPVVKKEERQNVNLTLP 151
Db 555 RKTAVEAATARAK-----ARKLEQQAQNAEP 580

RESULT 28
HBHA_MYCTU
ID HBHA_MYCTU STANDARD; PRT: 198 AA.
AC Q11142; O85733;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Heparin-binding hemagglutinin (Adhesin).
GN HBHA OR RV0475 OR MT0493 OR MTCY2069.01.
OS Mycobacterium tuberculosis, and
OS Mycobacterium bovis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae.
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773, 1765;
RN [1]
RP SEQUENCE FROM N.A., PARTIAL SEQUENCE, AND CHARACTERIZATION.
RC SPECIES=M.tuberculosis, and M.bovis;
RC STRAIN=H37Rv, H37Ra, and BCG / Paris 1173 P2;
RX MEDLINE=98445421; PubMed=9770536;
RA Menozzi F.D., Bischoff R., Fort E., Brennan M.J., Lochter C.;
RT "Molecular characterization of the mycobacterial heparin-binding
hemagglutinin, a mycobacterial adhesin.";

RL RP SEQUENCE FROM N.A.
RN SPECIES=M.tuberculosis; STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
Gordon S.V., Eiglmeyer K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sultun J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence.";
RL Nature 393:537-544(1998).
RN [3]
RP SEQUENCE FROM N.A.
RN SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 1-16, AND CHARACTERIZATION.
RN SPECIES=M.tuberculosis, and M.bovis;
RC STRAIN=H37Ra, and BCG / Paris 1173 P2;
RX MEDLINE=97188915; PubMed=9064359;
RA Menozzi F.D., Rouse J.H., Alavi M., Laude-Sharp M., Muller J.,
RA Bischoff R., Brennan M.J., Lochter C.;
RT "Identification of a heparin-binding hemagglutinin present in
Mycobacteria.";
RL J. Exp. Med. 184:993-1001(1996).
RN [5]
RP FUNCTION.
RC SPECIES=M.tuberculosis, and M.bovis;
RC STRAIN=103, and BCG;
RX MEDLINE=21342355; PubMed=11449276;
RA Pethe K., Alonso S., Biet F., Delogu G., Brennan M.J., Lochter C.,
RA Menozzi F.D.;
RT "The heparin-binding haemagglutinin of M. tuberculosis is required for
extrapulmonary dissemination.";
RL Nature 412:190-194(2001).
CC -!- FUNCTION: REQUIRED FOR EXTRAPULMONARY DISSEMINATION. MEDIATES
ADHERENCE TO EPITHELIAL CELLS BY BINDING TO SULFATED
GLYCOCONGUGATES PRESENT AT THE SURFACE OF THESE CELLS; BINDS
HEPARIN, DEXTRAN SULFATE, FUCOIDAN AND CHONDROITIN SULFATE.
PROMOTES HEMAGGLUTINATION OF ERYTHROCYTES OF CERTAIN HOST SPECIES.
CC -!- INDUCES MYCOBACTERIAL AGGREGATION.
CC -!- SUBCELLULAR LOCATION: SURFACE ASSOCIATED.
CC -!- DOMAIN: HEPARIN BINDING SEEMS TO REQUIRE THE C-TERMINAL
DOMAIN OF HBHA. PROGRESSIVE TRUNCATIONS FROM THE C-TERMINAL END
DIMINISH THE AFFINITY FOR HEPARIN.
CC -!- PTM: GLYCOSYLATED. GLYCOSYLATION MAY PROTECT THE PROTEIN FROM
PROTEOLYTIC DEGRADATION AND BE IMPORTANT FOR HEMAGGLUTINATION. IT
SUGGESTS THAT THE CARBOHYDRATE MOIETY MAY BE LOCATED WITHIN THE C-
TERMINAL DOMAIN OF HBHA.
CC -!- MISCELLANEOUS: SERUM FROM PATIENTS DIAGNOSED WITH ACTIVE
TUBERCULOSIS THAT HAD NOT BEEN VACCINATED CONTAINS ANTIBODIES THAT
RECOGNIZE HBHA, WHEREAS SERUM FROM HEALTHY INDIVIDUALS DOES NOT
CONTAIN ANY.
CC -!- SIMILARITY: STRONG, TO M.LEPRAE HBHA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
```



```
RT *Relationship of M protein genes in group A streptococci.*;
RL Proc. Natl. Acad. Sci. U.S.A. 82:1822-1826(1985).
CC -!- FUNCTION: THIS PROTEIN IS ONE OF THE DIFFERENT ANTIGENIC SEROTYPES
CC OF PROTEIN M. PROTEIN M IS CLOSELY ASSOCIATED WITH VIRULENCE OF
CC THE BACTERIUM AND CAN RENDER THE ORGANISM RESISTANT TO
CC PHAGOCYTOSIS.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein. Cell wall.
CC -!- SIMILARITY: TO OTHER M PROTEINS.
CC -!- SIMILARITY: TO OTHER STREPTOCOCCAL AND STAPHYLOCOCCAL PROTEINS
CC IN THE REGION OF THE MEMBRANE ANCHOR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M11338; AAA26920.1; -.
CC PIR; A26297; A26297.
CC InterPro; IPR001899; Gram_pos_anchor.
CC InterPro; IPR003345; M_repeat.
CC Pfam; PF00746; Gram_pos_anchor; 1.
CC Pfam; PF02370; M; 9.
CC PRINTS; PR00015; GP05ANCHOR.
CC PROSITE; PS00343; GRAM_POS_ANCHORING; 1.
CC Virulence; Phagocytosis; Cell wall; Repeat; Antigen; Transmembrane;
CC Coiled coil; Signal.
CC -----
CC SIGNAL 1 42
CC CHAIN 43 483 M PROTEIN, SEROTYPE 6.
CC DOMAIN 43 457 EXTRACELLULAR (POTENTIAL).
CC TRANSMEM 458 477 MEMBRANE ANCHOR.
CC DOMAIN 478 483 CYTOPLASMIC (POTENTIAL).
CC DOMAIN 69 138 10 X 7 AA TANDEM REPEATS.
CC DOMAIN 157 269 4.5 X 25 AA TANDEM REPEATS.
CC DOMAIN 279 347 TWO DIRECTLY REPEATED 27 AMINO ACID
CC BLOCKS SEPARATED BY 15 AMINO ACIDS.
CC DOMAIN 348 411 HYDROPHILIC.
CC DOMAIN 412 448 GLY/PRO-RICH (CELL WALL-SPANNING).
CC DOMAIN 449 454 CONSERVED IN GRAM-POSITIVE COCCI SURFACE
CC PROTEINS.
CC -----
CC QUERY MATCH 12.2%; Score 95; DB 1; Length 483;
CC Best Local Similarity 29.6%; Pred. No. 13;
CC Matches 48; Conservative 22; Mismatches 62; Indels 30; Gaps 9;
CC -----
CC QY 21 LGNAPEKALNRNEAIDELKKQATE-----DKATTATAEASD--ALEAL---A 66
CC DB 304 LANLTAELDKVKEEKQISDSRQGLRRDLDSAREAKKQVEKALEANSKLALEKLNKEL 363
CC QY 67 DOTDALQSEEAHVKKAD-NAASDAL-EALADQTDALQSEEAHVQSDNAASDAWE----- 119
CC DB 364 EESKKLTEKEAKLEQAKLEAKALKEQAKAEALAKLAGKA-SDSOTPDAPKGNKV 422
CC QY 120 --KAATPTALDVKKTKDPVVKVKEERQNVNTLPTTGESNP 159
CC DB 423 PKGQAPQA--GTPKNQNKAPMKETKQ-----LPSTGETANP 458
CC -----
CC RESULT 31
CC M12_STRPY STANDARD; PRT; 564 AA.
CC AC P19401;
CC DT 01-NOV-1990 (Rel. 16, Created)
CC DT 01-NOV-1990 (Rel. 16, Last sequence update)
CC DT 16-OCT-2001 (Rel. 40, Last annotation update)
CC DE M protein, serotype 12 precursor (Fragment).
CC GN EM12.
CC OS Streptococcus pyogenes.
CC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
CC Streptococcus.
```

```
OX NCBI_TaxID-1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CS24 / Serotype M12;
RX MEDLINE=88058777; PubMed=2445730;
RA Robbins J.C., Spanier J.G., Jones S.J., Simpson W.J., Cleary P.P.;
RT "Streptococcus pyogenes type 12 M protein gene regulation by upstream
RL J. Bacteriol. 169:5633-5640(1987).
CC -!- FUNCTION: THIS PROTEIN IS ONE OF THE DIFFERENT ANTIGENIC SEROTYPES
CC OF PROTEIN M. PROTEIN M IS CLOSELY ASSOCIATED WITH VIRULENCE OF
CC THE BACTERIUM AND CAN RENDER THE ORGANISM RESISTANT TO
CC PHAGOCYTOSIS.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein. Cell wall.
CC -!- SIMILARITY: TO OTHER M PROTEINS.
CC -!- SIMILARITY: TO OTHER STREPTOCOCCAL AND STAPHYLOCOCCAL PROTEINS
CC IN THE REGION OF THE MEMBRANE ANCHOR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M18269; AA88573.1; -.
CC PIR; A60115; A60115.
CC InterPro; IPR001899; Gram_pos_anchor.
CC InterPro; IPR003345; M_repeat.
CC Pfam; PF00746; Gram_pos_anchor; 1.
CC Pfam; PF02370; M; 9.
CC PROSITE; PS00343; GRAM_POS_ANCHORING; 1.
CC Virulence; Phagocytosis; Cell wall; Repeat; Antigen; Transmembrane;
CC Coiled coil; Signal.
CC -----
CC SIGNAL 1 41
CC CHAIN 42 >564 M PROTEIN, SEROTYPE 12.
CC DOMAIN 42 550 EXTRACELLULAR (POTENTIAL).
CC TRANSMEM 551 >564 MEMBRANE ANCHOR.
CC DOMAIN 44 505 COILED COIL (POTENTIAL).
CC DOMAIN 505 541 GLY/PRO-RICH (CELL WALL-SPANNING).
CC DOMAIN 542 547 CONSERVED IN GRAM-POSITIVE COCCI SURFACE
CC PROTEINS.
CC -----
CC NON_TER 564 564
CC SEQUENCE 564 AA; 62904 MW; 5F1549DACAA77B46 CRC64;
CC -----
CC QUERY MATCH 12.2%; Score 95; DB 1; Length 564;
CC Best Local Similarity 29.6%; Pred. No. 15;
CC Matches 48; Conservative 22; Mismatches 62; Indels 30; Gaps 9;
CC -----
CC QY 21 LGNAPEKALNRNEAIDELKKQATE-----DKATTATAEASD--ALEAL---A 66
CC DB 337 LANLTAELDKVKEEKQISDSRQGLRRDLDSAREAKKQVEKALEANSKLALEKLNKDL 456
CC QY 67 DOTDALQSEEAHVKKAD-NAASDAL-EALADQTDALQSEEAHVQSDNAASDAWE----- 119
CC DB 457 EESKKLTEKEAKLEQAKLEAKALKEQAKAEALAKLAGKA-SDSOTPDAPKGNKAV 515
CC QY 120 --KAATPTALDVKKTKDPVVKVKEERQNVNTLPTTGESNP 159
CC DB 516 PKGQAPQA--GTPKNQNKAPMKETKQ-----LPSTGETANP 551
CC -----
CC RESULT 32
CC HLES_DROME STANDARD; PRT; 1077 AA.
CC ID HLES_DROME
CC AC Q02308;
CC DT 01-FEB-1994 (Rel. 28, Created)
CC DT 01-FEB-1994 (Rel. 28, Last sequence update)
CC DT 01-FEB-1994 (Rel. 28, Last annotation update)
CC DE Hairless protein.
CC GN H.
CC OS Drosophila melanogaster (Fruit fly).
```



CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; X16421; CAA344442.1; ALT\_INIT.  
DR PIR; S05542; S05542.  
DR MEROPS; C40.001; -.  
DR InterPro; IPR000064; NLPC\_P60.  
DR Pfam; PF00877; NLPC\_P60; I.  
KW Signal; Cell wall.  
FT SIGNAL 1 27 POTENTIAL.  
FT CHAIN 28 516 P54 PROTEIN.  
FT ACT\_SITE 429 429 POTENTIAL.  
SQ SEQUENCE 516 AA; 54596 MW; 402ECAA439846D26 CRC64;

Query Match 12.0%; Score 93.5; DB 1; Length 516;  
Best Local Similarity 25.0%; Pred. No. 17;  
Matches 41; Conservative 29; Mismatches 79; Indels 15; Gaps 4;  
QY 2 DSPIEQPTIPNGGTLTNLIGNAPEKLALRN-----EERAIDELKKQAIEDKATTAT 54  
DB 167 ENAEKQELADNOALESQKGLLAKQADLVNLTSLAAEQATAEDKADLNKRKAEEA 226  
QY 55 EAASSDALEALADOTDALQSEEAAYVVKADNAASDALEALADOTDALQSEEAAYVQSDNAA 114  
DB 227 EQAIRREQARLAQA-----RQAAQEKAEKAREQAAQAQOAL-SSASTTTESSAA 281  
QY 115 SDAEKAAATPIALDVKYKDFPVVKKERQNVNLTPTGSESN 158  
DB 282 QSSSEESKAP---ESSTTEESTSTSTTNSSTGSSSTESS 322

RESULT 35  
ABPX\_YEAST  
ID ABPX\_YEAST STANDARD; PRT; 627 AA.  
AC Q08641; Q08644;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE Actin-binding protein ABP140.  
GN ABP140 OR YOR239W/YOR240W.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes.  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=S288C / FY1679;  
RX MEDLINE=97127829; PubMed=8972580;  
RA Boyer J., Michaux G., Fairhead C., Gaillon L., Dujon B.;  
RT "Sequence and analysis of a 26.9 kb fragment from chromosome XV of  
RT the yeast Saccharomyces cerevisiae.";  
RL Yeast 12:1575-1586(1996).  
RN [2]  
RP PARTIAL SEQUENCE FROM N.A., SEQUENCE OF 1-12; 38-72; 102-115; 244-286;  
RX 560-574 AND 598-608, FUNCTION, AND SUBCELLULAR LOCATION.  
RC STRAIN=BJ5457;  
RX MEDLINE=98127445; PubMed=9467951;  
RA Asakura T., Sasaki T., Nagano F., Satoh A., Obaishi H., Nishiohara H.,  
RA Imamura H., Hotta K., Tanaka K., Nakanishi H., Takai Y.;  
RT "Isolation and characterization of a novel actin filament-binding  
RT protein from Saccharomyces cerevisiae.";  
RL Oncogene 16:121-130(1998).

CC -1- FUNCTION: Binds F-actin and shows weak F-actin crosslinking  
CC activity.  
CC -1- SUBCELLULAR LOCATION: Cytoplasm and cortical cytoskeleton.  
CC -1- MISCELLANEOUS: A RIBOSOMAL FRAMESHIFT OCCURS BETWEEN THE CODONS  
CC FOR LEU-276 AND GLY-277.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; Z75147; CAA99460.1; ALT\_SEQ.  
DR EMBL; Z75147; CAA99461.1; ALT\_SEQ.  
DR SGD; S0005765; ABP140.  
DR InterPro; IPR001601; Meth-transf.  
DR InterPro; IPR000051; SAM\_bind.  
KW Actin-binding; Cytoskeleton; Ribosomal frameshift.  
FT INIT\_MET 0  
SQ SEQUENCE 627 AA; 71354 MW; D4E55F9485412F39 CRC64;

Query Match 12.0%; Score 93.5; DB 1; Length 627;  
Best Local Similarity 28.1%; Pred. No. 21;  
Matches 39; Conservative 23; Mismatches 60; Indels 17; Gaps 5;  
QY 27 KLALRNEPRAIDELKKQAIEDKATTATTAAS----SDAL-EALADQTDAL----- 72  
DB 97 KIYSGENADINVDFQYKEMENGTGAELASSVESDAIOEGVAETEGTIATPKOKENE 156  
QY 73 OSEEAAYVVKADNAASDALEALADOTDA--LQSEEAAYVQSDNAAADAEKAAATPIALDVK 130  
DB 157 KNDSEEEESANNASEPAEYQSSEEDADIEQSNKRETAENASQQA-NDGSTSTTTSKN 215  
QY 131 KTKDTKPVVKKERQNVNT 149  
DB 216 KKKKKKKKKKKRNGVNT 234

RESULT 36  
MOES\_MOUSE  
ID MOES\_MOUSE STANDARD; PRT; 576 AA.  
AC P26041;  
DT 01-MAY-1992 (Rel. 22, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Moesin (Membrane-organizing extension spike protein).  
GN MSN.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93055012; PubMed=1429901;  
RA Sato N., Funayama N., Nagafuchi A., Yonemura S., Tsukita S.,  
RA Tsukita S.;  
RT "A gene family consisting of ezrin, radixin and moesin. Its specific  
RT localization at actin filament/plasma membrane association sites.";  
RL J. Cell Sci. 103:131-143(1992).  
RN [2]  
RP SEQUENCE OF 11-576 FROM N.A.  
RX MEDLINE=92243764; PubMed=1573844;  
RA Furthmayr H., Lankes W.T., Amleu M.R.;  
RT "Moesin, a new cytoskeletal protein and constituent of filopodia: its  
RT role in cellular functions.";  
RL Kidney Int. 41:665-670(1992).  
CC -1- FUNCTION: PROBABLY INVOLVED IN CONNECTIONS OF MAJOR CYTOSKELETAL  
CC STRUCTURES TO THE PLASMA MEMBRANE.  
CC -1- SIMILARITY: CONTAINS 1 BAND 4.1-LIKE DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; S47577; AAA11762.1; -.  
DR EMBL; M86390; AAA39728.1; -.





RL Mol. Biol. Cell 7:1167-1180(1996).  
 RN [4]  
 RP SEQUENCE OF 18-1278 FROM N.A.  
 RX MEDLINE=99023772; PubMed=9805007;  
 RA Tokito M.K., Holzbaur E.L.F.;  
 RT "The genomic structure of DCTN1, a candidate gene for limb-girdle  
 muscular dystrophy";  
 RL Biochim. Biophys. Acta 1442:432-436(1998).  
 CC -!- FUNCTION: REQUIRED FOR THE CYTOPLASMIC DYNEIN-DRIVEN RETROGRADE  
 CC MOVEMENT OF VESICLES AND ORGANELLES ALONG MICROTUBULES. DYNEIN-  
 CC DYNACTIN INTERACTION IS A KEY COMPONENT OF THE MECHANISM OF AXONAL  
 CC TRANSPORT OF VESICLES AND ORGANELLES.  
 CC -!- SUBUNIT: LARGE MACROMOLECULAR COMPLEX OF AT LEAST 10 COMPONENTS.  
 CC P150(GLUED) BINDS DIRECTLY TO MICROTUBULES AND TO CYTOPLASMIC  
 CC DYNEIN.  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; P150 (SHOWN HERE) AND P135;  
 CC ARE PRODUCED BY ALTERNATIVE SPLICING.  
 CC -!- TISSUE SPECIFICITY: BRAIN.  
 CC -!- PTM: PHOSPHORYLATED.  
 CC -!- SIMILARITY: BELONGS TO THE DYNACTIN 150 KDA SUBUNIT FAMILY.  
 CC -!- SIMILARITY: CONTAINS 1 CAP-GLY DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF064205; AAD55811.1; JOINED.  
 DR EMBL; AF064203; AAD55811.1; JOINED.  
 DR EMBL; AF064204; AAD55811.1; JOINED.  
 DR EMBL; X98801; CA667333.1; JOINED.  
 DR EMBL; AF086947; AAD03694.1; JOINED.  
 DR EMBL; AF086927; AAD03694.1; JOINED.  
 DR EMBL; AF086928; AAD03694.1; JOINED.  
 DR EMBL; AF086929; AAD03694.1; JOINED.  
 DR EMBL; AF086930; AAD03694.1; JOINED.  
 DR EMBL; AF086931; AAD03694.1; JOINED.  
 DR EMBL; AF086932; AAD03694.1; JOINED.  
 DR EMBL; AF086933; AAD03694.1; JOINED.  
 DR EMBL; AF086934; AAD03694.1; JOINED.  
 DR EMBL; AF086935; AAD03694.1; JOINED.  
 DR EMBL; AF086936; AAD03694.1; JOINED.  
 DR EMBL; AF086937; AAD03694.1; JOINED.  
 DR EMBL; AF086938; AAD03694.1; JOINED.  
 DR EMBL; AF086939; AAD03694.1; JOINED.  
 DR EMBL; AF086940; AAD03694.1; JOINED.  
 DR EMBL; AF086941; AAD03694.1; JOINED.  
 DR EMBL; AF086942; AAD03694.1; JOINED.  
 DR EMBL; AF086943; AAD03694.1; JOINED.  
 DR EMBL; AF086944; AAD03694.1; JOINED.  
 DR EMBL; AF086945; AAD03694.1; JOINED.  
 DR EMBL; AF086946; AAD03694.1; JOINED.  
 DR MIM; 601143; .  
 DR InterPro; IPR000938; CAP-GLY.  
 DR Pfam; PF01302; CAP\_GLY; 1.  
 DR PROSITE; PS00845; CAP\_GLY\_1; 1.  
 DR PROSITE; PS0245; CAP\_GLY\_2; 1.  
 KW Motor protein; Microtubules; Dynein; Coiled coil; Cytoskeleton;  
 KW Alternative splicing; Phosphorylation.  
 FT DOMAIN 48 90 CAP-GLY.  
 FT DOMAIN 164 191 SER-RICH.  
 FT DOMAIN 213 547 COILED COIL (POTENTIAL).  
 FT DOMAIN 943 1049 COILED COIL (POTENTIAL).  
 FT DOMAIN 1182 1211 COILED COIL (POTENTIAL).  
 FT VARSPPLIC 1 131 MAQSKRHVYSTRPSCSRMSASAPLRVSGRVGVICKGHR  
 FT GTVAYVGLATLFGKVGIVLDEAKKNDGTVQGRKFTCD  
 FT ECHGIFVQSQIQFEDAGDTSPETPDSSASKVLKRGFTD  
 FT TTKTSKL -> MMRQ (IN ISOFORM P135).  
 FT VARIANT 163 163 A -> P.

FT CONFLICT 10 10 /FTID=VAR\_001373.  
 FT CONFLICT 132 138 S -> N (IN REF. 2 AND 3).  
 FT CONFLICT 712 712 MISSING (IN REF. 2 AND 3).  
 FT CONFLICT 712 712 D -> V (IN REF. 2 AND 3).  
 SQ SEQUENCE 1278 AA; 141694 MW; 6DCBA5E67B56E4BC CRC64;  
 Query Local 11.9%; Score 92.5; DB 1; Length 1278;  
 Best Local Similarity 28.3%; Pred. No. 52;  
 Matches 28; Conservative 21; Mismatches 39; Indels 11; Gaps 2;  
 OY 28 LALRNEERAIDELKKO-AIEDKEATTA-----TEAASSDALEALADOTDALQSEE 76  
 DB 943 LKLEDRTEVKEIKSLKIKGSEANVRLSLEKLDLSAAKADRIEKVQTRLEETQ 1002  
 OY 77 AAVVKNADNAASDALEALADOTDALQSEEAENVVQSDNAAS 115  
 DB 1003 ALLRKRKEFEETMDALQADIDQLEAKELQRLNSQS 1041  
 RESULT 39  
 HTR2\_HALN1 STANDARD; PRT; 763 AA.  
 AC Q9HP81;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Sensory rhodopsin II transducer (HTR-II) (Methyl-accepting phototaxis  
 DE protein II) (MPP-II).  
 GN HTR2 OR VNG1765G.  
 OS Halobacterium sp. (strain NRC-1).  
 OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;  
 OC Halobacterium.  
 OX NCBI\_TaxID=64091;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20504483; PubMed=11016950;  
 RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,  
 RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,  
 RA Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,  
 RA Leithauser B., Keller K., Cruz R., Danson M.J., Hough D.W., Dale H.,  
 RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.W., Jung K.-H.,  
 RA Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,  
 RA Alam M., Freitas T., Hou S., Daniels C., Dennis P.P., Omer A.D.,  
 RA Ebhardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.;  
 RT "Genome sequence of Halobacterium species NRC-1";  
 RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).  
 CC -!- FUNCTION: TRANSDUCES SIGNALS FROM THE PHOTOTAXIS RECEPTOR SENSORY  
 CC RHODOPSIN II (SR-II) TO THE FLAGELLAR MOTOR. RESPONDS TO LIGHT  
 CC CHANGES THROUGH THE VARIATION OF THE LEVEL OF METHYLATION (BY  
 CC SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO THE CHEMOTAXIS SENSORY TRANSDUCERS FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AE005080; AAG19989.1; .  
 DR InterPro; IPR004089; Chemotaxis\_transducer.  
 DR InterPro; IPR003660; HAMP.  
 DR Pfam; PF00672; HAMP; 1.  
 DR Pfam; PF00015; MCPsignal; 1.  
 DR SMART; SM00304; HAMP; 2.  
 DR SMART; SM00283; MA; 1.  
 KW Transducer; Photoreceptor; Transmembrane; Methylation;  
 KW Complete proteome.  
 FT INIT\_MET 0 0 BY SIMILARITY.  
 FT DOMAIN 1 15 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 16 36 POTENTIAL.  
 FT DOMAIN 37 277 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 278 297 POTENTIAL.  
FT DOMAIN 298 763 CYTOPLASMIC (POTENTIAL).  
SQ SEQUENCE 763 AA; 78911 MW; CF7A8FF04DFE309A CRC64;

Query Match 11.8%; Score 91.5; DB 1; Length 763;  
Best Local Similarity 29.1%; Pred. No. 35;  
Matches 37; Conservative 20; Mismatches 53; Indels 17; Gaps 5;  
QY 37 IDELKQKQAIEDKEATTATIEAASSDALEALADQTDALQSEEAAYVKADNAASDALEALADQ 96  
Db 504 VDDLSSAAEEVAETVASLADTAGQAASAV---DDGRQATEDAVETMDDVADDA-EAAADA 559  
QY 97 TDALQSEEAAYVQSDNAASDAWEKA---ATPIALDVKKT---KDTKPVVKKEERQNTL 150  
Db 560 MDALDSEADIGEIVDVIADIADQTNMLNALNASTEAARTGADGDFAVVADE-----V 612  
QY 151 PTTGEES 157  
Db 613 KTLAEEES 619

RESULT 40  
HTR2\_HALSA STANDARD; PRT; 764 AA.  
AC P71410;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Sensory rhodopsin II transducer (HTR-II) (Methyl-accepting phototaxis  
protein II) (MPP-II).  
GN HTR2 OR HTRII.  
OS Halobacterium salinarium.  
OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;  
OC Halobacterium.  
OX NCBI\_TaxID-2242;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=FLX15;  
RX MEDLINE=96323203; PubMed-8710852;  
RA Zhang W., Brooun A., Mueller M.M., Alam M.;  
RT "The primary structures of the Archaeon Halobacterium salinarium blue  
light receptor sensory rhodopsin II and its transducer, a methyl-  
accepting protein.";  
RL Proc. Natl. Acad. Sci. U.S.A. 93:8230-8235(1996).  
CC -1- FUNCTION: TRANSDUCES SIGNALS FROM THE PHOTOTAXIS RECEPTOR  
SENSORY RHODOPSIN II (SR-II) TO THE FLAGELLAR MOTOR. RESPONDS TO  
LIGHT CHANGES THROUGH THE VARIATION OF THE LEVEL OF METHYLATION.  
CC -1- SIMILARITY: BELONGS TO THE CHEMOTAXIS SENSORY TRANSDUCERS FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: U62676; AAC44369.1; -.  
DR HSSP: P02942; IQU7.  
DR InterPro: IPR004089; Chemotaxis\_transducer.  
DR InterPro: IPR003660; HAMP.  
DR Pfam: PF00672; HAMP; 1.  
DR Pfam: PF00015; MCPsignal; 1.  
DR SMART: SM00304; HAMP; 2.  
DR SMART: SM00283; MA; 1.  
KW Transducer; Photoreceptor; Transmembrane; Methylation.  
FT INIT\_MET 0 0 BY SIMILARITY.  
FT DOMAIN 1 15 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 16 36 POTENTIAL.  
FT DOMAIN 37 277 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 278 298 POTENTIAL.  
FT DOMAIN 299 764 CYTOPLASMIC (POTENTIAL).  
SQ SEQUENCE 764 AA; 79187 MW; 1E0D7B4E460FC588 CRC64;

Query Match 11.8%; Score 91.5; DB 1; Length 764;  
Best Local Similarity 29.1%; Pred. No. 35;  
Matches 37; Conservative 20; Mismatches 53; Indels 17; Gaps 5;  
QY 37 IDELKQKQAIEDKEATTATIEAASSDALEALADQTDALQSEEAAYVKADNAASDALEALADQ 96  
Db 505 VDDLSSAAEEVAETVASLADTAGQAASAV---DDGRQATEDAVETMDDVADDA-EAAADA 560  
QY 97 TDALQSEEAAYVQSDNAASDAWEKA---ATPIALDVKKT---KDTKPVVKKEERQNTL 150  
Db 561 MDALDSEADIGEIVDVIADIADQTNMLNALNASTEAARTGADGDFAVVADE-----V 613  
QY 151 PTTGEES 157  
Db 614 KTLAEEES 620  
Search completed: October 13, 2002, 04:46:23  
Job time : 27.9572 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2002, 04:38:53 ; Search time 43.3636 Seconds  
(without alignments)  
352.328 Million cell updates/sec

Title: US-09-847-539A-6  
 perfect score: 777  
 Sequence: 1 VDSPIEQPRIIPNGGTLNL.....KKEERONVNTLPTTGEESNP 159

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

```
Database :      PIR_71:*
1:  pir1:*
2:  pir2:*
3:  pir3:*
4:  pir4:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	188	24.2	448	2	A24496	IgG-binding protein G precursor
2	187	24.1	593	2	S00128	multiple ligand-binding protein
3	151.5	19.5	439	2	S67921	albumin-binding protein
4	138.5	17.8	323	2	A44801	S-antigen precursor
5	131.5	16.9	309	1	YACQN7	probable tail fiber
6	128.5	16.5	438	2	D90734	probable tail fiber
7	128.5	16.5	440	2	F85584	probable tail comp
8	122.5	15.8	1110	2	I51116	NF-180 - sea lamp
9	119.5	15.4	437	2	H90854	probable tail fiber
10	119.5	15.4	971	2	B90835	probable tail fiber
11	119.5	15.4	973	2	C85693	probable membrane
12	116	14.9	166	2	T48231	hypothetical protein
13	114.5	14.7	413	2	S53690	plasma protein recu
14	111	14.3	415	2	S35760	fcrA protein precu
15	111	14.3	2055	2	T31110	extracellular matr
16	110.5	14.2	401	1	QX9P1L	hypothetical prote
17	110	14.2	490	2	I41024	colicin 10 - Esche
18	110	14.2	890	2	S22452	surface exclusion
19	109.5	14.1	407	2	G90907	probable tail fiber
20	109.5	14.1	437	2	E90968	probable tail fiber
21	109.5	14.1	439	2	E85816	probable tail fiber
22	109	14.0	394	2	F90725	membrane spanning
23	109	14.0	394	2	G85576	membrane spanning
24	108.5	14.0	271	2	G90898	probable tail fiber
25	108.5	14.0	421	2	JV0057	tolA protein - Esc
26	108	13.9	159	2	T10699	hypothetical prote
27	108	13.9	405	2	A33939	Fc gamma (19G) rec
28	108	13.9	1096	2	T13802	centrosome-associ
29	107	13.8	1358	2	S52536	fcr15 protein -

## ALIGNMENTS

RESULT 1

A24496

IGG-binding protein - Streptococcus sp. (group G)

C:Species: Streptococcus sp.

C:Date: 17-Sep-1987 #sequence\_revision 17-Sep-1987 #text\_change 17-Mar-2000

C:Accession: A24496; A39041

R:Fahnestock, S.R.; Alexander, P.; Nagle, J.; Filipula, D.

J. Bacteriol. 167, 870-880, 1986

A:Title: Gene for an immunoglobulin-binding protein from a group G Streptococcus.

A:Reference number: A24496; MUID:86304178

A:Accession: A24496

A:Molecule type: DNA

A:Residues: 1-448 <FAH>

A:Cross-references: GB:MI3825; NID:gl53822; PIDN:AAA03664.1; PID:gl53823

R:Sjoeberring, U.; Bjoerck, L.; Kaster, W.

J. Biol. Chem. 266, 399-405, 1991

A:Title: Streptococcal protein G. Gene structure and protein binding properties.

A:Reference number: A39041; MUID:91093154

A:Accession: A39041

A>Status: preliminary

A:Molecule type: protein

A:Residues: 34-42, 'N', 45-48; 62-76; 186-200 <SJO>

RESULT 2  
S001x28  
protein G precursor - Streptococcus sp. (Streptococcus G148)  
N:Alternate names: albumin-binding protein; cell wall-bound protein  
C:Species: Streptococcus sp.  
A:Variety: Streptococcus G148  
C:Date: 30-Jun-1989; sequence-revision 30-Jun-1989 #text\_change 19-  
C:Accession: S00128; 827404; A26314



	Query Match	15.8%	Score 122.5;	DB 2;	Length 1110;
	Best Local Similarity	29.4%;	Pred. No. 1;		
	Matches	37;	Conservative	28;	Mismatches 50; Indels 11; Gaps 5;
Qy	38	DELKQAATEDKENTATAEAASSDLEALAD--QTDLQSEEAUVVKADNAA--SDALEALA	94		
		:  :        :  :  :      :			
Dd	705	DEABEEVYKEEVT-----KSDAEEAFAEEAAEKSEEEAAAEAKDEABEEAAVEAV	758		
Qy	95	DOTDALOSE-RAEVVQSDNAASDAWEKAATPIALDVKKTKDTKPVVKKERQNVTLPPT	153		
		:  :        :  :  :  :  :  :  :  :			
Dd	759	EETGAATEEAEEAKEASDDDEKPEEEVYKESEAPVAPEAKKAPFPKPAKPKAKVES-PTS	817		
Ov	154	GEESNP	159		









DNA Res. 8, 11-22, 2001

R; Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, S.; Sasawara, N.; Yasunaga, T.; Kubara, S.; Shiba, T.; Hattori, M.; Shinagawa, A.

RESULT 24  
G90898  
probable: tail fiber protein [imported] - *Escherichia coli* (strain 0157:H7, substrain C), Species: *Escherichia coli*  
C.Date: 18-Jul-2001 #sequence-revision 18-Jul-2001 #text-change 18-Jul-2001  
C.Accession: G90898  
R.Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.  
Gisawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.





```

Db      144 SAANNPTSAGALESARQMAESAAAAAQSEASSSSSAAAAQAKSESSQSAAEHLSRKT 203
QY      114 ASDAMEKAAPIALDVKKTKNDKPVYKKEER-----QNVNTLPT 152
        | | | | | | | | | | | | | | | | | | | | | |
Db      204 AESAAGNNAARDATTATKEARSESASQASQESRTAAEAVNRPT 248

```

### RESULT 32

G85631  
hypothetical protein Z1382 [imported] - Escherichia coli (strain O157:H7, substrain EDL933)  
C:Species: Escherichia coli  
C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001  
C:Accession: G85631  
R:Perera, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
Miller, L.; Grobbeck, E.J.; Davis, N.W.; Lim, A.; Dimantista, E.; Potamoultis, K.; Apodaca,  
Nature 409, 523-533, 2001  
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
A:Reference number: AB5480; MUID:21074935; PMID:11206551  
A:Accession: G85631  
A:Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-375 <STO>  
A:Cross-references: CB:AE005174; NID:g12514226; PIDN:ANG55515.1; GSPDB:GN00145; UWGP:Z13  
A:Experimental source: strain O157:H7, substrain EDL933  
?:Genetics:  
?:Gene: Z1382

### RESULT 33

GB6887  
 Probable tail fiber protein GP37 - *Escherichia coli*  
 C:Species: *Escherichia coli*  
 C:Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 21-Jul-2000  
 C:Accession: G64887; T09189  
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co  
 A.: Rose, D.J.; Mau, B.; Shao, Y.  
 S:Science 277, 1453-1462, 1997  
 A:Title: The complete genome sequence of *Escherichia coli* K-12.  
 A:Reference number: A64720; MUID:97426617  
 A:Accession: G64887  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-1122 <BAT>  
 A:Cross-references: GB:AE000234; GB:U00096; NID:g1787633; PIDN:AACT4454.1; PID:g1787636;  
 A:Experimental source: strain K-12, substrain MO1655  
 R:Alta, H.; Baba, T.; Fujita, K.; Hayashi, K.; Inada, T.; Isono, K.; Itoh, T.; Kasai, H.  
 I.: Motomura, K.; Nakade, S.; Nakamura, Y.; Nishimoto, H.; Nishio, Y.; Oshima, T.; Saito,  
 moto, Y.; Horikuchi, T.  
 DNA Res. 3, 363-377, 1996  
 A:Title: A 570-kb DNA sequence of the *Escherichia coli* K-12 genome corresponding to the  
 A:Reference number: Z16603; MUID:97251357  
 A:Accession: T09189  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 3-1122 <AIB>  
 A:Cross-references: GB:AE000234; GB:U00096; NID:g1787633; PIDN:AACT4454.1; PID:g1787636;

Query Match	13.6%;	Score 105.5;	DB 2;	Length 1122;
Best Local Similarity	31.5%;	Pred. No. 14;		
Matches	41;	Conservative	18;	Mismatches 58;
			Indels	13;
			Gaps	4;

```

Oy 67 DOTLAL0EEFAAVVKADNAASDALLEALADQ-TDIAL0EEAEVVOSDNAASAMWAKATPI 125
Db 146 HAADAAADSAARAASRISAGOMASSASQASSACTASTKATEA----SKSMAAAESSKMAAT 201
Oy 126 ALDYKTKTKDT 135
Db 202 SAGAARKTSET 211

```

## RESULT 34

557834  
 fera protein precursor - Streptococcus pyogenes  
 C:Species: Streptococcus pyogenes  
 C:Date: 28-Oct-1995 #sequence\_revision 03-Nov-1995 #text\_change 26-Aug-1999  
 C:Accession: S57834  
 R:Boyle, M.D.P.; Hawlitzky, J.; Raeder, R.; Podbelski, A.  
 Infect. Immun. 62, 1336-1347, 1994  
 A:Title: Analysis of genes encoding two unique type IIIa immunoglobulin G-binding proteins  
 A:Reference number: S57834, MUID:94178942  
 A:Accession: S57834  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-387 <BOY>  
 A:Cross-references: EMBL:X72932; NID:g507128; PIDN:CA51436.1; PID:g507129  
 S:Superfamily: M5 protein

QY	97	TDAILOSBEAEVYQSDDNAADAMEKAAPPIALDYVKTKTDKTVKKERONVN-----	148
Dd	300	L-ACAAEELIAVLAKDKDMSG----AOKP---DTKPGNKEVPTRPSOTRTNTNKASMAQT	350
QY	149	-TLPTTGEE-SNP 159   ::     :	
Dd	351	RELTPSTGETTNP 363   ::     :	

### RESULT 35

A46173  
MRP4 protein - Streptococcus sp. (group A)  
C:Species: Streptococcus sp.  
C:Date: 21-Sep-1993 #sequence\_revision 25-Apr-1997 #text\_change 30-May-1998  
C:Accession: A46173  
R:O'Toole, P.; Stenberg, L.; Rissler, M.; Landahl, G.  
Proc. Natl. Acad. Sci. U.S.A. 89, 8661-8665, 1992  
A:Title: Two major classes in the M protein family in group A streptococci  
A:Reference number: A46173; MUID:92409576  
A:Contents: group A  
A:Accession: A46173  
A:Status: preliminary  
A:Molecule type: nucleic acid  
A:Residues: 1-388 <017>  
A:Note: sequence extracted from NCBI backbone (NCBIN:114063, NCBIPI:114064)  
C:Superfamily: M5 protein

Query Match 13.5%; Score 105; DB 2; Length 388;  
Best Local Similarity 26.9%; Pred. No. 4.5;  
Matches 52; Conservative 26; Mismatches 67; Indels 48; Gaps 10;

QY 6 EOPRIINGCTLTNLGNA-----PEKIALRNERAIDEL---KKQAIED 47  
DB 181 EIAKIQSEAAFTLEMLGSAKRELTELQAKIDPTATAEKAKLESQVLTLEMLGSAKRELTD 240  
QY 48 KEATTAIEAASSDLEA----LADQTDALQSEEA---AVYADANAASDALE---ALADQ 96  
DB 241 LQALDANAENEKELQSOATLEKOLENTKELADLQAKLATATQEKLEAEKALKLEQ 300  
QY 97 TDALQSEAEVYVQSDNADAMEKAATPIALDVKRTKTPVYKKEERQNVN----- 148  
DB 301 L-AAQAELEAKLKADKXASG-----AQKP---DTKPKKEVPTRESQRTNTNKKAPMAQTK 351  
QY 149 -TLPTGEE-SNP 159  
DB 352 RQLPSTGEETNP 364

RESULT 36  
S42574  
streptococcal surface protein - Streptococcus dysgalactiae  
C:Species: Streptococcus dysgalactiae  
C:Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 19-May-2000  
C:Accession: S42574  
R:Jonsen, H.; Mueller, H.P.  
Eur. J. Biochem. 220, 819-826, 1994  
A:Title: The type-III PC receptor from Streptococcus dysgalactiae is also an alpha(2)-ma  
A:Reference number: S42574; MUID:94192673  
A:Accession: S42574  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-664 <JON>  
A:Cross-references: EMBL:Z29666; NID:9470220; PIDN:CAA82764.1; PID:9470221  
C:Superfamily: M5 protein

Query Match 13.5%; Score 105; DB 2; Length 664;  
Best Local Similarity 23.9%; Pred. No. 8.3;  
Matches 54; Conservative 21; Mismatches 75; Indels 76; Gaps 9;

QY 1 VDSPIEOP-----RIIPNGCTLTNLGNAPKALRNERAIDELKQAIED----- 47  
DB 423 IDAPELPALTYKLYKGNTFSS---GETTKAV--DAETAEKAFKQVANENGVCSESY 477  
QY 48 KEATTAIEAASSDLEAALADQTDALQSEEA-VKADNADALEALADQTDALQSEEA-E 106  
DB 478 DDATKTFVTVEKPAVIDAPELTPALTYKLYI---NGKTLKGETTKAVDAETAEKAFK 533  
QY 107 VVQSDNADAMW-----EK-----AATPIADV 129  
DB 534 QYAVENGVGVYDDATKTFVTVEWTEVPGDAFTEPEKEPASIPVLPAPPIAKDD 593  
QY 130 KKTDTKTPV-----VKKEERQNVLTPTGSESNP 159  
DB 594 AKKDDTKKVDTKKEDAKKPEAKKEAKKATILPTGEGSNP 639

RESULT 37  
C90769  
probable tail fiber protein (imported) - Escherichia coli (strain 0157:H7, substrain RIM  
C:Species: Escherichia coli  
C:Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 18-Jul-2001  
C:Accession: C90769  
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.  
gasawara, N.; Yasunaga, T.; Kuhnara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
DNA Res. 8, 11-22, 2001  
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli 0157:H7 and genc  
A:Reference number: A99629; MUID:21156231; PMID:11258796  
A:Accession: C90769  
A:Status: preliminary  
A:Molecule type: DNA

A:Residues: 1-439 <HAY>  
A:Cross-references: GB:BA000007; PIDN:BA834546.1; PID:913360583; GSPDB:GN00154  
A:Experimental source: strain 0157:H7, substrain RIMD 0509952  
C:Genetics:  
A:Gene: Ecs1123

Query Match 13.4%; Score 104.5; DB 2; Length 439;  
Best Local Similarity 26.4%; Pred. No. 5.6;  
Matches 47; Conservative 20; Mismatches 76; Indels 35; Gaps 4;

QY 15 GTLTNLGNAPD-----KIALRNERAIDELKQAIEDK----- 49  
DB 86 GTLNDFLGMSDEDVREPALRFELWVEEAARHAEKAKNGAETSARNAGISASQAE 145  
QY 50 ----ATTATIEAASSDLEAALADQTDALQSEEA-VKADNADALEALADQTDALQSEEA 105  
DB 146 SAANADTSAGDASASARQALESAAAQSEASSSSSSAAQKASQSSQSAAEELSKRT 205  
QY 106 EVVQSDNADAMEKAATPIALDVKRTKTPVYKKEERQNVNLTPTT---GESNP 159  
DB 206 AESAAGNARDA--TTATKEKRESSESQSAEQSRIAAEEAVNRIPTVYGPGRKGP 261

RESULT 38  
S72375  
surface exclusion protein sepl precursor - Enterococcus faecalis plasmid P01  
C:Species: Enterococcus faecalis  
C:Date: 12-Feb-1998 #sequence\_revision 13-Mar-1998 #text\_change 15-Oct-1999  
C:Accession: S72375  
R:Hirt, H.; Wirth, R.; Muscholl, A.  
Mol. Gen. Genet. 252, 640-647, 1996  
A:Title: Comparative analysis of 18 sex pheromone plasmids from Enterococcus faecalis  
A:Reference number: S72375; MUID:97074879  
A:Accession: S72375  
A:Molecule type: DNA  
A:Residues: 1-843 <HIR>  
A:Cross-references: EMBL:X96976; NID:91272652; PIDN:CAA65662.1; PID:e235489; PID:9127  
A:Experimental source: strain OG1X  
C:Genetics:  
A:Gene: sepl  
A:Genome: Plasmid P01  
F:1-30/Domain: signal sequence #status predicted <SIG>  
F:31-843/Product: surface exclusion protein #status predicted <MAT>

Query Match 13.4%; Score 104.5; DB 2; Length 843;  
Best Local Similarity 28.3%; Pred. No. 12;  
Matches 39; Conservative 27; Mismatches 61; Indels 11; Gaps 5;

QY 7 QPRIIPNGCTLTNLGNAPKALRNERAIDELKQAIEDKQATTAIEAASSDLEALA 66  
DB 32 QPK-----TEFNASSEPTVYKATQTTEDALTEKQOYIE-KQAIVDQKQOAVDPAKREK 84  
QY 67 DQTD-ALQSEEA-VKADNADALEALADQTDALQSEEA-EVQSDNADAMEKAATPI 125  
DB 85 DTIDQSVKDOQAVVNDKDALVQSOQAVTDQ-QAVVDPAKRV--DEATPAIEKAKQV 141  
QY 126 ALDYAKTKTDTKPYAKKEE 143  
DB 142 ATDTQAVDDQOKVVEQAO 159

RESULT 39  
S33441  
EF protein - Streptococcus suis  
C:Species: Streptococcus suis  
C:Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 15-Oct-1999  
C:Accession: S33441  
R:Smith, H.E.; Reek, F.H.; Veicht, U.; Gielkens, A.L.J.; Smits, M.A.  
submitted to the EMBL Data Library, May 1993  
A:Description: Repeats in an extracellular protein of wek-pathogenic strains are abse  
A:Reference number: S33441  
A:Accession: S33441  
A:Status: preliminary

A: Molecule type: DNA  
A: Residues: 1-1822 <SMI>  
A: Cross-references: EMBL:X71880; NID:g2298031; PIDN:CAA50714.1; PID:g2298032

Query Match 13.4%; Score 104.5; DB 2; Length 1822;  
Best Local Similarity 24.5%; Pred. No. 28;  
Matches 35; Conservative 32; Mismatches 53; Indels 23; Gaps 5;

OY 16 TLTNLGNAPKELRNNEERAIDELKKQAT--EDKEATTAIEAASSDALEALADQTD-AL 72  
DB 1214 SLDTGTKEARDAVELAKDK----ELAKEARTTEEBEATKIVEKLAEDTTRAKAIEDNPRLSD 1269  
OY 73 QSEEAAYKADNAASDALEALADQTDALQSEEAAYQSDNAASDAWEKAATPIALDVKKT 132  
DB 1270 EDKQAEIKKLTDAVAKTLATIRDNADK-RTQAEKKAQA-----LADLEKA 1313  
OY 133 KDTKPYVKKKEERQNVNTLPPTGE 155  
DB 1314 KETQKTIADRAIDRLTLVKDGE 1336

## RESULT 40

G41662  
130k surface exclusion protein Sec10 precursor - Enterococcus faecalis plasmid pCF10  
C: Species: Enterococcus faecalis  
C: Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 15-Oct-1999  
C: Accession: G41662  
R: Kuo, S.M.; Olmsted, S.B.; Vikarins, A.S.; Gallo, J.C.; Dunny, G.M.  
J: Bacteriol. 173, 7650-7664, 1991  
A: Title: Molecular and genetic analysis of a region of plasmid pCF10 containing positive  
terococcus faecalis.  
A: Reference number: A41662; MUID:92041679  
A: Accession: G41662  
A: Status: preliminary  
A: Molecule type: DNA  
A: Residues: 1-891 <KAO>  
A: Cross-references: GB:M64978; NID:g150552; PIDN:AAA65847.1; PID:g150554  
C: Genetics:  
A: Genome: plasmid

Query Match 13.4%; Score 104; DB 2; Length 891;

Best Local Similarity 26.9%; Pred. No. 14;

Matches 42; Conservative 27; Mismatches 73; Indels 14; Gaps 6;

OY 7 OPRIPNGGTLTLGNAPKELRNNEERAIDELKKQATIEDKREATTATIEAASSDALEALA 66  
DB 32 QRTTPNSST-----EQPTVAKATQTTTEQALTE-KQQVTEKQAIYDOKQOVADTKAKEK 84  
OY 67 DQTDALQSEEAAYKADNAASD-ALFALADQTDALQSEEAAYQSDNAASDAWEKAATPI 125  
DB 85 DAIDQSVKDAQAVVDQNKDALDQSQQAVTDQ-QAVVDEAKKV--DEATPSAIEKAKEQV 141  
OY 126 ALDVKKTQDKPYVKKKE--RONVNTLPPTGEEEN 158  
DB 142 ATDTQAVDEQKVVDAQQTDVNOQAVVDEKAKETN 177

Search completed: October 13, 2002, 04:49:16  
Job time : 47.3636 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OW protein - nucleic search, using frame\_plus.p2n model

Run on: October 13, 2002, 04:54:53 : Search time 51.016 Seconds  
(without alignments)  
765.558 Million cell updates/sec

Title: US-09-847-539A-6  
Perfect score: 777  
Sequence: 1 VDSPIRQPRIPNGTFLTNL.....KKEPQNVNLTPTGESNP 159

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 38353 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame\_plus.p2n.model -DEV=xih  
-O=/cgn2\_1/USPTO.spool/US09847539/rnatc\_10102002\_093106\_5020/app\_query.fasta.1.526  
-DB=Issued\_Patents\_NA -QEMT=fastap -SUFFIX=std.rni -MINMATCH=0.1 -LOOPEXT=0  
-LIST=45 -DOCALLIG=200 -THR.SCORE=PCT -THR.MAX=100 -THR.MIN=0 -ALIGN=40  
-MODE=LOCAL -OUTFMT=PCT -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09847539\_6CGL1\_1\_13\_4rnatc\_10102002\_093106\_5020 -NCPU=6 -ICPU=3  
-NO\_XLPHY -NO\_MMAB -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120  
-WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: Issued\_Patents\_NA.\*  
2: /cgn2\_6/ptodata/2/ina/5A.COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/5B.COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6A.COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/6B.COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/6CTUS.COMB.seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	114.5	14.7	1555	3 US-08-669-408B-9	Sequence 9, Appli
2	105.5	13.6	989	4 US-09-121-517B-5	Sequence 1, Appli
3	105	13.5	2526	4 US-08-669-408B-1	Sequence 1, Appli
4	104.5	13.4	6744	1 US-08-119-125A-2	Sequence 2, Appli
5	101	13.0	2706	2 US-08-630-822A-61	Sequence 61, Appli
6	101	13.0	2706	2 US-09-005-069-61	Sequence 1, Appli
7	97.5	12.5	4248	3 US-08-678-614-1	Sequence 224, App
8	96.5	12.4	1886	4 US-08-936-165A-224	Sequence 1, Appli
9	96	12.4	824	1 US-08-158-353-1	Sequence 1, Appli
10	94.5	12.2	937	4 US-08-821-872-1	Sequence 5, Appli
11	93	12.0	1332	2 US-08-795-475-5	Sequence 9, Appli
12	92.5	11.9	1419	2 US-08-216-894-9	

13	92.5	11.9	1419	4 US-09-115-746-9	Sequence 9, Appli
14	92.5	11.9	3279	4 US-08-446-137B-1	Sequence 1, Appli
15	92.5	11.9	4376	1 US-08-119-125A-1	Sequence 1, Appli
16	91.5	11.8	3256	2 US-08-968-751-3	Sequence 3, Appli
17	91	11.7	3622	2 US-08-591-079-7	Sequence 7, Appli
18	91	11.7	5393	2 US-08-591-079-9	Sequence 9, Appli
19	90.5	11.6	1548	2 US-08-762-106-5	Sequence 5, Appli
20	90.5	11.6	1548	4 US-09-320-774-5	Sequence 5, Appli
21	90.5	11.6	1581	2 US-08-752-106-6	Sequence 6, Appli
22	90.5	11.6	1581	2 US-08-752-106-6	Sequence 6, Appli
23	90	11.6	943	1 US-08-106-981-4	Sequence 1, Appli
24	90	11.6	2085	1 US-08-072-070-1	Sequence 1, Appli
25	90	11.6	2085	1 US-08-465-746-1	Sequence 1, Appli
26	90	11.6	2085	1 US-08-214-164-1	Sequence 1, Appli
27	90	11.6	2085	1 US-08-469-434-1	Sequence 1, Appli
28	90	11.6	2085	1 US-08-214-222-1	Sequence 1, Appli
29	90	11.6	2085	2 US-08-467-852A-1	Sequence 1, Appli
30	90	11.6	2085	2 US-08-468-718-1	Sequence 1, Appli
31	90	11.6	2085	2 US-08-246-636-1	Sequence 1, Appli
32	90	11.6	2085	2 US-08-247-491A-1	Sequence 1, Appli
33	90	11.6	2085	2 US-08-319-795-1	Sequence 1, Appli
34	90	11.6	2085	2 US-08-468-985-1	Sequence 1, Appli
35	90	11.6	2086	3 US-08-312-949-1	Sequence 1, Appli
36	90	11.6	2086	3 US-08-446-201-2	Sequence 1, Appli
37	88.5	11.4	918	3 US-08-937-271-9	Sequence 9, Appli
38	88.5	11.4	1560	1 US-07-813-584A-1	Sequence 1, Appli
39	88.5	11.4	1560	1 US-08-330-515-1	Sequence 1, Appli
40	88.5	11.4	1771	2 US-08-533-669A-7	Sequence 7, Appli
41	88.5	11.4	1771	2 US-08-511-872-1	Sequence 1, Appli
42	88.5	11.4	2241	2 US-08-785-431-3	Sequence 3, Appli
43	88.5	11.4	2241	4 US-09-205-048-3	Sequence 3, Appli
44	88.5	11.4	2367	2 US-08-785-431-1	Sequence 1, Appli
45	88.5	11.4	2367	4 US-09-205-048-1	Sequence 1, Appli

#### ALIGNMENTS

RESULT 1  
Sequence 9, Application US/08669408B  
Patent No. 6100055  
GENERAL INFORMATION:  
APPLICANT: JONSSON, Bengt  
APPLICANT: LINDERBERG, Martin  
APPLICANT: MUELLER, Hans-Peter  
APPLICANT: RANTAMAKI, Liisa K.  
TITLE OF INVENTION: METHOD AND MEANS FOR PRODUCING  
PLASMAPROTEINASE INHIBITOR-BINDING PROTEINS  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/669,408B  
FILING DATE: 03-JUL-1996  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/SE94/00826  
FILING DATE: 06-SEP-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: SE 9302855-3  
FILING DATE: 06-SEP-1993  
ATTORNEY/AGENT INFORMATION:



Sequence 1, Application US/08669408B  
Patent No. 6100055  
GENERAL INFORMATION:  
APPLICANT: GUS, Bengt  
APPLICANT: JONSSON, Hans  
APPLICANT: LINDBERG, Martin  
APPLICANT: MUELLER, Hans-Peter  
APPLICANT: RANTAMAKI, Liisa K.  
TITLE OF INVENTION: METHOD AND MEANS FOR PRODUCING  
TITLE OF INVENTION: PLASMAPROTEINASE INHIBITOR-BINDING PROTEINS  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/669,408B  
FILING DATE: 03-JUL-1996  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/SE94/00826  
FILING DATE: 06-SEP-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: SE 9302855-3  
FILING DATE: 06-SEP-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 61743/102  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2526 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 506..2497  
NAME/KEY: mat\_peptide  
LOCATION: 506..2497  
US-08-669-408B-1  
Alignment Scores:  
Pred. No.: 0.00638 Length: 2526  
Score: 105.00 Matches: 54  
Percent Similarity: 33.19% Conservative: 21  
Best Local Similarity: 23.89% Mismatches: 75  
Query Match: 13.51% Indels: 76  
DB: 3 Gaps: 9  
US-09-847-539A-6 (1-159) x US-08-669-408B-1 (1-2526)  
QY 1 ValAspSerProIleGluInPro-----ArgIleIleProAsnGlyGly 15  
DB 1772 AFTGAGCGACCTGAGACTACCGCTTGACTACTACAACTATTTGTTAAAGTAAAC 1831  
QY 16 ThrLeuThrAsnLeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArg 35  
DB 1832 ACTTTCCTCA-----GGCGAAACAACACTACTAAAGCAGTA-----GACGCAAGAACT 1876

QY 36 AlaIleAspGluLeuLysLysGluAlaIleGluAsp----- 47  
DB 1877 GCAGAAAAAGCCTTCAAACATACGTACGAAAAAGCGTGTTCAGGTGATGCTTAT 1936  
QY 48 LysGluAlaThrThrAlaIleGluAlaIleSerSerAspAlaLeuGluAlaLeuAlaAsp 67  
DB 1937 GACCATGCACTAAACCTTTACAGTTACTGAAAAACCGACGATGTGACGCACTGAA 1996  
QY 68 GlnPThrAspAlaLeuGlnSerGluGluAlaIleValValLysAlaAspAsnAlaIleSer 87  
DB 1997 TTAAACACCGATTCGACCACTACCAAACTGTTATC-----AATGTTAAACA 2044  
QY 88 AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaIleGlu 106  
DB 2045 TTGAAGGCGAAGACACTACCTAAAGCACTACAGCCAGAACTGCAGAAAAAGCCTTCAA 2104  
QY 107 ValValGlnSerAspAsnAlaIleSerAspAlaIleTrp----- 118  
DB 2105 CAATPACGCTAACGAAAAACGGTGTGATGCTGTGACTTACGATGATGCGACTAAGACC 2164  
QY 119 -----GluLys 120  
DB 2165 TTACAGTAACTGAAGTTACTGAAAGTCTCGTGATGCAACCACTGAACCAAGAAC 2224  
QY 121 -----AlaAlaThrProIleAlaLeuAspVal 129  
DB 2225 CCAGAACCAAGTATCCCTCTGTTCCCTTAACCTCTGCAACTCCCAATTCCTTAAGATGAC 2284  
QY 130 LysLysThrLysAspThrLysProVal----- 138  
DB 2285 GCTAAGAAAGACACTACTAAGAAAGTGCATCTAAGAAAGAACGCTAAACCAAGAA 2344  
QY 139 -----ValLysLysGluLysArgGlnAsnValAsnThrLeuProThrThr 153  
DB 2345 GCTAAGAAAGAAAGAAAGCTAAGAAAGAAAGAAAGCTAAGAAAGCTTCTTCAACT 2404  
QY 154 GlyGluGlnSerAsnPro 159  
DB 2405 GGTGAAGAGAACCAACCA 2422  
RESULT 4  
US-08-119-125A-2  
Sequence 2, Application US/08119125A  
Patent No. 5610011  
GENERAL INFORMATION:  
APPLICANT: SMITH, Hilda Elizabeth  
TITLE OF INVENTION: DNA Sequences which code for Virulence  
TITLE OF INVENTION: Characteristics of Streptococcus suis and the use thereof for the di  
TITLE OF INVENTION: antibodies derived therefrom and the use thereof for the di  
TITLE OF INVENTION: protection against infection by S. suis in mammals, Includi  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Central Diergeneeskundig Instituut  
STREET: Edelhertweg 15  
CITY: PH Leijstad  
STATE: The Netherlands  
COUNTRY: The Netherlands  
ZIP: NL-8219  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage  
COMPUTER: IBM compatible  
OPERATING SYSTEM: MS-DOS v.6.0  
SOFTWARE: Wordperfect v. 6.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/119,125A  
FILING DATE: 20-SEP-1993  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/NL92/00054  
FILING DATE: 19-MAR-1992  
APPLICATION NUMBER: NL 9100510  
FILING DATE: 21-MAR-1991

ATTORNEY/AGENT INFORMATION:  
NAME: Hanab, Anthony H.  
REGISTRATION NUMBER: 26275  
REFERENCE/DOCKET NUMBER: SMITHHE119125  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (203) 838-8589  
TELEFAX: (203) 838-8794  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 6744 base pairs  
TYPE: Nucleic acid with corresponding amino acids  
STRANDEDNESS: single stranded  
TOPOLOGY: linear  
MOLECULE TYPE: genomic DNA  
ORIGINAL SOURCE:  
ORGANISM: Streptococcus suis type II (pathogenic)  
FEATURE:  
OTHER INFORMATION: Extracellular factor related protein (EF\*) gene  
FEATURE:  
NAME/KEY: promoter -35 region  
LOCATION: bp 66 to 71  
FEATURE:  
NAME/KEY: promoter -10 region  
LOCATION: bp 89 to 94  
FEATURE:  
NAME/KEY: promoter -35 region  
LOCATION: bp 153 to 158  
FEATURE:  
NAME/KEY: promoter -10 region  
LOCATION: bp 176 to 181  
FEATURE:  
NAME/KEY: ribosome binding site  
LOCATION: bp 350 to 356  
FEATURE:  
NAME/KEY: signal peptide  
LOCATION: bp 361 to 498  
FEATURE:  
NAME/KEY: start of repetitive units RI-RI1  
LOCATION: bp 2869, 3097, 3292, 3520, 4087, 4381, 4609, 4837,  
LOCATION: 5065, 5293, 5521:  
FEATURE:  
NAME/KEY: start of repetitive Asn-Pro-Asn-Leu sequences  
LOCATION: bp 2932, 3160, 3355, 3583, 4150, 4444, 4672, 4900,  
LOCATION: 5128, 5356, 5584:  
FEATURE:  
NAME/KEY: dyad symmetry regions  
LOCATION: from bp 6554 to 6566 and from bp 6571 to 6583  
FEATURE:  
NAME/KEY: dyad symmetry regions  
LOCATION: from bp 6611 to 6625 and from bp 6631 to 6644  
US-08-119-125A-2  
Alignment Scores:  
Pred. No.: 0.0262 Length: 6744  
Score: 104.50 Matches: 35  
Percent Similarity: 46.858 Conservative: 32  
Best Local Similarity: 24.488 Mismatches: 53  
Query Match: 13.458 Indels: 23  
DB: 1 Gaps: 5  
US-09-847-539A-6 (1-159) x US-08-119-125A-2 (1-6744)  
QY 16 ThleuThrAsnLeuLeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluIuArg 35  
DB 4000 ACCTTGACAGACAGCAGTAAGACGATGCAATTCGCTGAAGATTAATAA--- 4056  
QY 36 AlaIleAspGluLeuLysGlnAlaIle-----GluAspLysGluAlaThrThrAla 53  
DB 4057 -----GAATTACTTAAGCAAGCAATCCGAACAGAACAAAGAAAGCTACTAAATA 4107  
QY 54 ILeGluAlaIleAspAlaLeuAlaLeuAlaLeuAlaLeuAlaLeuAlaLeuAlaLeu 72  
DB 4108 GTAGAAACAACTTCAGACAGATACGCCAAAGCTATCGACAGCAATCAAACTTGCGAT 4167

QY 73 GlnSerGluGluAlaAlaValValLysAlaAspAsnAlaAspAlaLeuGluAla 92  
DB 4168 GAACATTAAGCAGCGAAATTAATAAACCTACTGACGCTGGCGAAAACTTTGCAAC 4227  
QY 93 LeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaGluValValGlnSerAspAsn 112  
DB 4228 ATTGCTGACATGCGATTAAC---CTAGCGCAAGACGAAAGCAAAAGCTCAAGCC----- 4278  
QY 113 AlaIleAspAlaLeuThrProGluAlaLeuAspValLysThr 132  
DB 4279 -----CTAGCAGATCTTGAAAAAGCT 4299  
QY 133 LysAspThrLysProValValLysLysGluGluArgGlnAsnValAsnThrLeuProThr 152  
DB 4300 AAAGAAACACAGAAATTCGATTAAGCGCATTGATAGCTTACTATCTGTGAAA 4359  
QY 153 ThrLysGlu 155  
DB 4360 GATGTGAG 4368  
RESULT 5  
US-08-630-822A-61  
Sequence 61, Application US/08630822A  
Patent No. 5840695  
GENERAL INFORMATION:  
APPLICANT: FRANK, GLENN R.  
APPLICANT: HUNTER, SHIRLEY WU  
APPLICANT: MALLERFELS, LYNDY  
TITLE OF INVENTION: NOVEL ECTOPARASITE SALIVA PROTEINS  
TITLE OF INVENTION: AND APPARATUS TO COLLECT SUCH PROTEINS  
NUMBER OF SEQUENCES: 107  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sheridan Ross P.C.  
STREET: 1700 Lincoln Street, Suite 3500  
CITY: Denver  
STATE: Colorado  
COUNTRY: U.S.A.  
ZIP: 80203  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/630,822A  
FILING DATE: 11-Apr-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: CONNELL, GARY J.  
REGISTRATION NUMBER: 32,020  
REFERENCE/DOCKET NUMBER: 2618-17-C3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 863-9700  
TELEFAX: (303) 863-0223  
INFORMATION FOR SEQ ID NO: 61:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2706 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 5..2706  
US-08-630-822A-61  
Alignment Scores:  
Pred. No.: 0.0201 Length: 2706  
Score: 101.00 Matches: 41  
Percent Similarity: 46.308 Conservative: 34  
Best Local Similarity: 25.318 Mismatches: 65  
Query Match: 13.004 Indels: 22

DB: 2 Gaps: 5  
US-09-847-539A-6 (1-159) x US-08-630-822A-61 (1-2706)  
QY 17 LeuThAsnLeuGlYAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArgAla 36  
DB 1478 TTAGAAAAACAATTAATGCTTTACAGAAAA-----TTGAAGGCTGAATTAGACAC 1531  
QY 37 IleAspGluLeuLysGlnAlaIleGluAspLysGluAlaThrThrAlaIleGluAla 56  
DB 1532 AATCAGAAATTAAAGAGCAAGCTGTGAGCTTAGAGTTGCTCAGTCTGACTGACAA 1591  
QY 57 AlaSerSer-----AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGln 73  
DB 1592 CTGAAATTAATGAATTACAGAACTATGCAAGGTTTACAAACAAAGAGATGCTTACAA 1651  
QY 74 SerGluGluAlaAlaVal-----LysAlaAspAsnAlaAla 86  
DB 1652 CAAGAAAGTAGCATCTCTCCAAAGCAAACTTTCTCAAGAGAGAGCTCTAGATCACAAGCT 1711  
QY 87 SerAsp-----AlaLeuGluAlaLeuAlaAspGlnThrAspAla 99  
DB 1712 TCTGATATGCAATAGACTAGAGCAAAATTCAGGCTCTCATATTGAACTGAGACAT 1771  
QY 100 LeuGlnSerGluGluAlaGluValValGlnSerAspAsnAlaAlaSerAsp----- 116  
DB 1772 GTCAGAAATTTGTAGACAAAGTTACCCAAAGACAGACAACTATTGGAAAGATATCA 1831  
QY 117 AlaTrpGluLysAlaAlaThrProIleAlaLeuAspValLysLysThrLysAspThrLys 136  
DB 1832 ACATTGAGAAAGAAATGCTCTCTAGAAATTAGAAATTGAAAGACACAAACAAATAT 1891  
QY 137 ProValValLysLysGluGluArgGlnAsnValAsnThrLeuProThrThrGlyGluGlu 156  
DB 1892 GAGCAAGAGCTCAAGCACATCGCAAAATGAAAAATCAAGACTGCTCAATAAAGAA 1951  
QY 157 SerAsn 158  
DB 1952 GCAAAAT 1957  
RESULT 6  
US-09-005-069-61  
Sequence 61, Application US/09005069  
Patent No. 5932470  
GENERAL INFORMATION:  
APPLICANT: FRANK, GLENN R.  
APPLICANT: HUNTER, SHIRLEY WU  
TITLE OF INVENTION: NOVEL ECTOPARASITE SALIVA PROTEINS  
TITLE OF INVENTION: AND APPARATUS TO COLLECT SUCH PROTEINS  
NUMBER OF SEQUENCES: 107  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sheridan Ross P.C.  
STREET: 1700 Lincoln Street, Suite 3500  
CITY: Denver  
STATE: Colorado  
COUNTRY: U.S.A.  
ZIP: 80203  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/005,069  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/630,822  
FILING DATE: 11-APR-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: CONNELL, GARY J.  
REGISTRATION NUMBER: 32,020

REFERENCE/DOCKET NUMBER: 2618-17-C3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 863-9700  
TELEFAX: (303) 863-0223  
INFORMATION FOR SEQ ID NO: 61:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2706 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 5..2706  
US-09-005-069-61  
Alignment Scores:  
Pred. No.: 0.0201 Length: 2706  
Score: 101.00 Matches: 41  
Percent Similarity: 46.30% Conservative: 34  
Best Local Similarity: 25.31% Mismatches: 65  
Query Match: 13.00% Indels: 22  
Gaps: 5  
US-09-847-539A-6 (1-159) x US-09-005-069-61 (1-2706)  
QY 17 LeuThAsnLeuGlYAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArgAla 36  
DB 1478 TTAGAAAAACAATTAATGCTTTACAGAAAA-----TTGAAGGCTGAATTAGACAC 1531  
QY 37 IleAspGluLeuLysGlnAlaIleGluAspLysGluAlaThrThrAlaIleGluAla 56  
DB 1532 AATCAGAAATTAAAGAGCAAGCTGTGAGCTTAGAGTTGCTCAGTCTGACTGACAA 1591  
QY 57 AlaSerSer-----AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGln 73  
DB 1592 CTGAAATTAATGAATTACAGAACTATGCAAGGTTTACAAACAAAGAGATGCTTACAA 1651  
QY 74 SerGluGluAlaAlaVal-----LysAlaAspAsnAlaAla 86  
DB 1652 CAAGAAAGTAGCATCTCTCCAAAGCAAACTTTCTCAAGAGAGAGCTCTAGATCACAAGCT 1711  
QY 87 SerAsp-----AlaLeuGluAlaLeuAlaAspGlnThrAspAla 99  
DB 1712 TCTGATATGCAATAGACTAGAGCAAAATTCAGGCTCTCATATTGAACTGAGACAT 1771  
QY 100 LeuGlnSerGluGluAlaGluValValGlnSerAspAsnAlaAlaSerAsp----- 116  
DB 1772 GTCAGAAATTTGTAGACAAAGTTACCCAAAGACAGACAACTATTGGAAAGATATCA 1831  
QY 117 AlaTrpGluLysAlaAlaThrProIleAlaLeuAspValLysLysThrLysAspThrLys 136  
DB 1832 ACATTGAGAAAGAAATGCTCTCTAGAAATTGAAATGAAAGCAACAAACAAATAT 1891  
QY 137 ProValValLysLysGluGluArgGlnAsnValAsnThrLeuProThrThrGlyGluGlu 156  
DB 1892 GAGCAAGAGCTCAAGCACATCGCAAAATGAAAAATCAAGACTGCTCAATAAAGAA 1951  
QY 157 SerAsn 158  
DB 1952 GCAAAAT 1957  
RESULT 7  
US-08-678-614-1  
Sequence 1, Application US/08678614  
Patent No. 6013507  
GENERAL INFORMATION:  
APPLICANT: Delencastre, Herminia  
APPLICANT: Tomas, Alexander  
TITLE OF INVENTION: AUXILIARY GENES AND PROTEINS OF  
TITLE OF INVENTION: METHICILLIN RESISTANT BACTERIA AND ANTAGONISTS THEREOF  
NUMBER OF SEQUENCES: 1  
CORRESPONDENCE ADDRESS:

ADDRESSEE: David A. Jackson, Esq.  
STREET: 411 Hackensack Ave, Continental Plaza, 4th  
FLOOR  
CITY: Hackensack  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07601

COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/678,614  
FILING DATE: 10-JUL-1996  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Jackson Esq., David A.  
REGISTRATION NUMBER: 26,742  
REFERENCE/DOCKET NUMBER: 600-1-139  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-487-5800  
TELEFAX: 201-343-1684  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4248 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Staphylococcus aureus  
STRAIN: RUSA 266  
FEATURE:  
NAME/KEY: ORF  
LOCATION: 154..1410  
NAME/KEY: ORF  
LOCATION: 1497..3500  
US-08-678-614-1

Alignment Scores:  
Pred. No.: 0.0915 Length: 4248  
Score: 97.50 Matches: 34  
Percent Similarity: 39.46% Conservative: 24  
Best Local Similarity: 23.13% Mismatches: 62  
Query Match: 12.55% Indels: 27  
DB: 3 Gaps: 3

US-09-847-539A-6 (1-159) x US-08-678-614-1 (1-4248)

OY 37 IIEAaPGLuLeuLySLyGLuAlaIleGLuAsPlyGLuAlaThrThra----- 53  
|||  
DB 2310 ATTGGTAAATTGAAACAAAGCAATTAAAGATTGATGACGCAACAAATGCACAA 2369  
|||  
OY 54 IIEGLuAlaLeuSerSerAspAlaLeuGLuAlaLeuAlaSerAspAlaLeuGLu 73  
|||  
DB 2370 GTAGAACGCATTAAACAAAGCAATGATTAATCAACATACCTCTACACAA 2429  
|||  
OY 74 SerGLuAlaAlaValValValAlaSpaAlaAlaSerAspAlaLeuGLuAlaLeu 93  
|||  
DB 2430 CCTTAAGCAGCGCTCTTCAGCAATTTGACGAAGTTGTCAGCAACAATGATCAACA 2489  
|||  
OY 94 AlaaSPGLnThraSPAlaLeuGLnSerGLuAlaGLuValValGLnSerAspAla 113  
|||  
DB 2490 CCTTAAATCTCTATACCAACAAATGAGAGAGAGCTATTGAAGCTATTATGCA 2549  
|||  
OY 114 AlaaSerAspAlaTrpGLuIuVala-----AlaThrProIleAlaLeuAspValLys 131  
|||  
DB 2550 GCTTAAGATTCTCGTCTTAAGCAATTGAAAGCAGACACACATGCAAGATTAGAAAG 2609  
|||

OY 132 ThrLys----- 133  
|||  
DB 2610 GTTAAACGAGAAATCTCAAAATTCGAAATATTACTGACTACGCAACAAATG 2669  
|||  
OY 134 -----AspThrLysProValValLysLysGLuGLuArgGLaAsnValAsnThr 149  
|||  
DB 2670 GATGCCATTAAATGAAGTTAAAGCTGACACACGCTACAGAAAGCTCAAAATCTACAGTT 2729  
|||  
OY 150 LeuProThrThrGLyGLuGLu 156  
|||  
DB 2730 TCAAATGCAACAAATGAGAA 2750  
|||

RESULT 8  
US-08-936-165A-224/C  
Sequence 224, Application US/08936165A  
Patent No. 6348582

GENERAL INFORMATION:  
APPLICANT: Black, Michael  
APPLICANT: Burnham, Martin  
APPLICANT: Hodgson, John  
APPLICANT: Knowles, David  
APPLICANT: Lonetto, Michael  
APPLICANT: Nicholas, Richard  
APPLICANT: Pratt, Julie  
APPLICANT: Reichard, Richard  
APPLICANT: Rosenberg, Martin  
APPLICANT: Ward, Judith  
TITLE OF INVENTION: No. 6348582el Prokaryotic Polynucleotides,  
TITLE OF INVENTION: Polypeptides and Their Uses  
NUMBER OF SEQUENCES: 534  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham Corporation  
STREET: 709 Swedeland Road  
CITY: King of Prussia  
STATE: PA  
COUNTRY: USA  
ZIP: 19406-0939

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/936,165A  
FILING DATE: 24-SEP-1997  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/027,032  
FILING DATE: 24-SEP-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Gimm, Edward R  
REGISTRATION NUMBER: 38,891  
REFERENCE/DOCKET NUMBER: P50549  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 610-270-4478  
TELEFAX: 610-270-5090  
TELEX:  
INFORMATION FOR SEQ ID NO: 224:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1886 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: Genomic DNA  
US-08-936-165A-224

Alignment Scores:  
Pred. No.: 0.0414 Length: 1886  
Score: 96.50 Matches: 31  
Percent Similarity: 47.50% Conservative: 26  
Best Local Similarity: 25.83% Mismatches: 54  
Query Match: 12.42% Indels: 9  
DB: 4 Gaps: 3

```

Query Match: 12.36% Indels: 30
DB: 1 Gaps: 8

US-09-847-539A-6 (1-159) x US-08-158-353-1 (1-824)

QY 14 G1yG1yThrLeuThrAsnLeuLeuGlyAsnAlaProGluLys-----LeuAla 29
   ||| ||||| ::| |||||
Db 300 GGGCGCTAATCTTACACGAATTAAGTAAAAAATTACAGAACTCAACGCAGTTCTCGCC 359
QY 30 LeuArgAsnGluLys-----ArgAlaIleAspGluLeuLysGlnAlaIleGlu 46
   ::::: ||| ::||| |||||
Db 360 GTCAAAAGAGTTGAACCTGCTTGTCATCTATAGATGAAGTGTCAAGAAAGATTGCGG 419
QY 47 AspLysGlnAlaThrThrAlaIleGlnAlaIleSerAsp-----AlaLeuGlnAla 64
   ::| ||| ::| ||| ::::: |||
Db 420 AATTGTATAGCCCAAAATGGTTTAAATGGCGGTCTAAATCAAAACGAGCATTTGTTACG 479
QY 65 -----LeuAlaAspGlnThrAspAlaLeuGlnSerGluLys 76
   ::|||::: ||| |||||::: |||
Db 480 GGAGCCTACGTAATATCAACCTCAATATAGCAGAAAAATTAGATGCTTAAAAAATTCAAGA 539
QY 77 AlaAlaValAlaLysAlaAspAsnAlaAlaSerAspAlaLeuGlnAlaLeuAlaAspGln 96
   ||| ::| ||| ||||| ::|
Db 540 GAATTAAGAAGAAAAAATTGAA-----GATGCTAAAAAATGTAAACAAGCATTT 587
QY 97 ThrAspAlaLeuGlnSerGluGlnAlaGluValAlaGlnSerAspAsnAlaIleSerAsp 116
   ||||| |||||::| |||||::: |||||
Db 588 ACTATATACTAAATAAAAGTACTCATGCGGAGAACTCGGTATGCGAAATGGAGCTGATGAT 647
QY 117 AlaTrpGluLysAlaAlaThrProIleAlaLeuAspValLysThrLysAspThrLys 136
   ||| ||||| ||| ::||| ||||| |||||
Db 648 GCTATATCAAAAGCGGCT-----ATTATAAAACAATGCTACTAA 689
QY 137 ProValAlaLysLysGlu--GluArg-----GlnAsnValAsnThrLeuProThrThr 153
   ::||| ||||| ::||| |||
Db 690 GATAAAGGCTGCTCAAGCCTTGAAGAATTATTGAATCATGTAATAAACTTTGCAAAAGCA 749
QY 154 GlyGluGluSer 157
   ::|||:::
Db 750 GCTCAAGAAACA 761

RESULT 10
US-08-821-872-1
/ Sequence 1, Application US/08821872A
/ Patent No. 6204004
/ GENERAL INFORMATION:
/ APPLICANT: Jarels, James B
/ APPLICANT: Jarels, Karen
/ TITLE OF INVENTION: Immunodiagnostic Test for Enterohemorrhagic Escherichia
/ FILE REFERENCE: Seq. ID No. 6204004 Ref: 1419JK
/ CURRENT APPLICATION NUMBER: US/08/821,872A
/ CURRENT FILING DATE: 1997-03-21
/ NUMBER OF SEQ ID NOS: 2
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 1
/ LENGTH: 937
/ TYPE: DNA
/ ORGANISM: Escherichia coli
US-08-821-872-1

Alignment Scores:
Pred. No.: 0.0282 Length: 937
Score: 94.50 Matches: 46
Percent Similarity: 42.31% Conservative: 20
Best Local Similarity: 29.49% Mismatches: 67
Query Match: 12.16% Indels: 23
DB: 4 Gaps: 6

US-09-847-539A-6 (1-159) x US-08-821-872-1 (1-937)

QY 14 GlyGlyThrLeuThrAsnLeuLeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGlu 33
   ||||| ::|||::: ||||| |||||

```







QY 133 LysAspThrLysProValIValLysLysGluGluArgGln 145  
Db 1072 GCTGAGCCAGCAGAGGTTCGCCGAGCGAGAGCAGAGC 1110  
RESULT 14  
US-08-446-137B-1  
Sequence 1, Application US/08446137B  
Patent No. 6162903  
GENERAL INFORMATION:  
APPLICANT: Trowern, Angus R.  
APPLICANT: Atkinson, Anthony  
APPLICANT: Murphy, Jonathan P.  
APPLICANT: Laurence, Oliver S.  
APPLICANT: Dugdaley, Clive J.  
TITLE OF INVENTION: IMMUNOGLOBULIN BINDING PROTEINS DERIVED  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: SEED and BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/446,137B  
FILING DATE: 22-MAY-1995  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Mcmasters, David D.  
REGISTRATION NUMBER: 33,963  
REFERENCE/DOCKET NUMBER: 100084.406  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3279 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: N-terminal  
ORIGINAL SOURCE:  
ORGANISM: Peptostreptococcus asaccharolyticus  
STRAIN: 1018  
FEATURE:  
NAME/KEY: exon  
LOCATION: 103..3186  
IDENTIFICATION METHOD: experimental  
OTHER INFORMATION: /codon\_start= 280  
OTHER INFORMATION: /product= "mature protein L"  
OTHER INFORMATION: /evidence= EXPERIMENTAL  
OTHER INFORMATION: /number= 1  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 103..3186  
OTHER INFORMATION: /codon\_start= 103  
OTHER INFORMATION: /product= "Immature protein L"  
FEATURE:  
NAME/KEY: mat\_peptide  
LOCATION: 280..3183  
OTHER INFORMATION: /codon\_start= 280  
OTHER INFORMATION: /product= "mature protein L"  
FEATURE:  
NAME/KEY: misc\_signal

LOCATION: 208..279  
US-08-446-137B-1  
Alignment Scores:  
Pred. No.: 0.245  
Score: 92.50  
Percent Similarity: 41.40%  
Best Local Similarity: 26.75%  
Query Match: 11.90%  
DB: 4  
Gaps: 6  
US-09-847-539A-6 (1-159) x US-08-446-137B-1 (1-3279)  
QY 26 GluLysLeuAlaLeuArgGln-----GluArgAlaIleAspGluLeuLysGln 43  
Db 778 GAAAAATTAGCAGCAGCAAAAGAACGAAAGCAAAACATATGATGATTTAAACCTTA 837  
QY 44 AlaIleGluAspLysGlu-----AlaThrThrAlaIleGluAla 56  
Db 838 GCAGACAAAACAAAGATTAGCTAGAGAGATATGATTGACGCTACTGATGATGCA 897  
QY 57 -----AlaSerSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAla 71  
Db 898 ATCAATGATATCTGCTAGCTAGCAGCAGATGTAATGGAAGAAAGCTATCTGAAAAAGAAACA 957  
QY 72 LeuGlnSerGluGluAlaAlaValValLysAlaAspAsnAlaAlaSerAspAlaLeuGlu 91  
Db 958 CCAGAACCAAGAAAGAAAGAACTTACATCAACGTAATCTTTCGACATGGA----- 1011  
QY 92 AlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaGluValValGlnSerAsp 111  
Db 1012 -----AGCACACAAATACAGAAATTCGAAAGAACATTCGCAAAACCATCATCAGATGCT 1065  
QY 112 AsnAlaAlaSerAspAlaThrProGluLysAlaIleThrProIleAlaLeuAspVal----- 129  
Db 1066 TACGCTTACGCGAGATGCTTTAAAGAAACAAACGAGAAATATACGTATACGTTGACAGAT 1125  
QY 130 -----LysLysThrLysAspThrLysProVal 138  
Db 1126 AAGCCTTAACCTTTAAATTTAAATTCGCTGTAATAAAAGAAACCAAGCAAGCA----- 1182  
QY 139 ValLysLysGluGluArgGlnAsnValAsnThrLeuProThrThrGlyGlu 155  
Db 1183 ---AAAGAGAAAGTTAACAATCAAGTTAACTTTCGAGATGGAAG 1230  
RESULT 15  
US-08-119-125A-1  
Sequence 1, Application US/08119125A  
Patent No. 5610011  
GENERAL INFORMATION:  
APPLICANT: SMITH, Hilida Elizabeth  
APPLICANT: VECCHI, Uri  
TITLE OF INVENTION: DNA Sequences which code for Virulence  
TITLE OF INVENTION: Characteristics of Streptococcus suis and parts thereof, po  
TITLE OF INVENTION: antibodies derived therefrom and the use thereof for the di  
TITLE OF INVENTION: protection against infection by S. suis in mammals, includi  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Central Diergeneeskundig Instituut  
STREET: Edelhertweg 15  
CITY: PH Delftstad  
STATE:  
COUNTRY: The Netherlands  
ZIP: NL-8219  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage  
COMPUTER: IBM compatible  
OPERATING SYSTEM: MS-DOS v. 6.0  
SOFTWARE: Wordperfect v. 6.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/119,125A  
FILING DATE: 20-SEP-1993  
CLASSIFICATION: 514

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/NL92/00054  
FILING DATE: 19-MAR-1992  
APPLICATION NUMBER: NL 9100510  
FILING DATE: 21-MAR-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Handal, Anthony H.  
REGISTRATION NUMBER: 26275  
REFERENCE/DOCKET NUMBER: SMITHHE119125  
TELEPHONE: (203) 838-8589  
TELEFAX: (203) 838-8794  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4376 base pairs  
TYPE: Nucleic acid with corresponding amino acids  
STRANDEDNESS: single stranded  
TOPOLOGY: linear  
MOLECULE TYPE: genomic DNA  
ORIGINAL SOURCE:  
ORGANISM: Streptococcus suis type II (pathogenic)  
FEATURE:  
OTHER INFORMATION: Extracellular protein factor (EF) gene  
NAME/KEY: promoter -35 region  
LOCATION: bp 66 to 71  
FEATURE:  
NAME/KEY: promoter -10 region  
LOCATION: bp 89 to 94  
FEATURE:  
NAME/KEY: promoter -35 region  
LOCATION: bp 153 to 158  
FEATURE:  
NAME/KEY: promoter -10 region  
LOCATION: bp 176 to 181  
FEATURE:  
NAME/KEY: ribosome binding site  
LOCATION: bp 350 to 356  
FEATURE:  
NAME/KEY: signal peptide  
LOCATION: bp 361 to 498  
FEATURE:  
NAME/KEY: mature peptide  
LOCATION: bp 499 to 2890  
FEATURE:  
NAME/KEY: dyad symmetry regions  
LOCATION: from bp 4186 to 4198 and from bp 4203 to 4215  
FEATURE:  
NAME/KEY: dyad symmetry regions  
LOCATION: from bp 4243 to 4257 and from bp 4263 to 4276  
US-08-119-125A-1  
Alignment Scores:  
Pred. No.: 0.357 Length: 4376  
Score: 92.50 Matches: 38  
Percent Similarity: 45.45% Conservative: 27  
Best Local Similarity: 26.57% Mismatches: 45  
Query Match: 11.90% Indels: 34  
Gaps: 6  
DB: 1  
US-09-847-539a-6 (1-159) x US-08-119-125A-1 (1-4376)  
QY 34 GUAAGAlaIleaspGluLeuLysGlnAlaIleGluAspLysGluAlaThrThraAla 53  
Db 2842 GAAAGGCTATTACAGTCGACGCG-CAATCGCGACGAGCAAA-----GCG 2888  
QY 54 IleGluAlaIleAspSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGln 73  
Db 2889 GTAGCTCAATCGCCCAAGATGCTTACGCA-----GCGAACA 2930  
QY 74 SerGluGluAlaIleAlaValIleAlaAspAlaIleAspAlaLeuGluAlaLeu 93  
Db 2931 GATGCTTAAGATTAAGATTCGTAAGAAATCGACGCTGCTAAGTCACGATTCAGCGCAAT 2990

QY 94 AlaAspGlnThrAspAlaLeuGlnSerGluGluAlaGluValValGlnSerAsp---Asn 112  
Db 2991 CCNAACCTTGACAGTCGACGAGAAAGAAATCAAGTCAAGAACCGTAGTCAGATGCTAA 3050  
QY 113 AlaIleAspAlaIleAspAlaIleGluAlaIleAlaIleProIle----- 125  
Db 3051 GCTCGACAGATGCAATGATGCTTCAACAAAGTCACATCGAAGCGCAATCGCAGACGAC 3110  
QY 126 -----AlaLeuAspValLysLysThrLysAspThrLys 136  
Db 3111 AAGCGCTAGGCGCCATCGCCAAAGACATTCCTGATCCGCGAAA---CAAGATGCTAAG 3167  
QY 137 ProValIleLysLysGluGluAlaGlnAsnValAsnThrLeuProThrGlnGluGlu 156  
Db 3168 AACAGATTCCTTAAGAGGAGAAATCCGCTAAGTCAGTCAAT-----GAC 3212  
QY 157 SerAsnPro 159  
Db 3213 TCCAAATCCG 3221  
RESULT 16  
US-08-968-751-3  
Sequence 3, Application US/08968751  
Patent No. 5948643  
GENERAL INFORMATION:  
APPLICANT: Rubinfield, Bonnie  
APPLICANT: Polakis, Paul G.  
APPLICANT: Ligenfelter, Carol  
APPLICANT: Vuong, Terilyn T.  
TITLE OF INVENTION: MODULATORS OF BRCA1 ACTIVITY  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: ONYX Pharmaceuticals, Inc.  
STREET: 3031 Research Drive  
CITY: Richmond  
STATE: CA  
COUNTRY: USA  
ZIP: 94806  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/968,751  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Glotta, Gregory  
REGISTRATION NUMBER: 32,028  
REFERENCE/DOCKET NUMBER: ONYX1024 GG  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (510) 262-8710  
TELEFAX: (510) 222-9758  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3256 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 34..2541  
US-08-968-751-3  
Alignment Scores:  
Pred. No.: 0.316 Length: 3256  
Score: 91.50 Matches: 37  
Percent Similarity: 42.14% Conservative: 30  
Best Local Similarity: 23.27% Mismatches: 73  
Query Match: 11.78% Indels: 19







Score: 90.50 Matches: 47  
Percent Similarity: 39.88% Conservative: 22  
Best Local Similarity: 27.17% Mismatches: 41  
Query Match: 11.65% Indels: 63  
DB: 2 Gaps: 8

US-09-847-539A-6 (1-159) x US-08-762-106-6 (1-1581)

QY 35 ArgAlaIleAspGluLeuLysGlnAla----- 44  
DB 925 CGCGTGTGGACAGCTGTGGAGAGCGCGCGCTGTGCAGCCAAATCGAGCACC 984  
QY 45 IleGluAspLysGluAlaThrThra----- 53  
DB 985 CTGGCGGACAAAGCCCGCCAGCAGCGCCAGCGTGCAGCGCTTGGAGCGGACGCC 1044  
QY 54 -----IleGluAla----- 56  
DB 1045 AAGATGAGCGCGAGCTGCAGAGGAGAGCTGAGCGCGTGGCGGCGCCACGTTT 1104  
QY 57 -----AlaSerSerAspAlaLeuGluAlaLeuAlaAsp 67  
DB 1105 GTGCTGCGCGAGCTCAAGCCCGCGCGCGCGCGCGCTGCGCGAGCGCGCGCG 1164  
QY 68 GlnThrAspAlaLeuGlnSerGluAlaAlaValValys----- 81  
DB 1165 GAGCTGACGCGG---CAGCGCGAGAGAGCGCCCAACGCCAAGTGAGAGCGGACAGCGC 1221  
QY 82 -----AlaAspAsnAlaIleSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAla 99  
DB 1222 GAGCGCGCGAGAGAGCGCGCGCGCGCGCG---GAGCGCGCGCGGAGAGAGCAGAGCGC 1278  
QY 100 LeuGlnSerGluGluAlaGlnValValGlnSerAspAsnAlaIleSerAspAlaTrpGlu 119  
DB 1279 CTGCTGAGAGGTGGCGCC-----ACGCGCGCGCGGAGAGCGGAGAGAG 1323  
QY 120 LysAlaIleThrProIleAlaLeuAspValLysThrLysAspThrLysProVal--- 138  
DB 1324 CGCGCGAGAGAGCGCGCGCGCGCGCGCGCGCTGCGCGAGTGCGCGTGTGAC 1383  
QY 139 ValLysLysGluGluArgGlnAsnValAsnThrLeuPro 151  
DB 1384 GTGAGGCTGAGTGCCCAAGCGGTGAGAGCTGTGCC 1422

RESULT 22  
US-09-320-774-6  
Sequence 6, Application US/09320774  
Patent No. 6265545  
GENERAL INFORMATION:  
APPLICANT: Javrik, Jonathan W.  
TITLE OF INVENTION: READING FRAME INDEPENDENT EPITOPE  
TITLE OF INVENTION: TAGGING  
NUMBER OF SEQUENCES: 47  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Harris Brotman  
STREET: 202 Coast Blvd., Suite 111  
CITY: La Jolla  
STATE: California  
COUNTRY: US  
ZIP: 92037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/320,774  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/762,106  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:

NAME: Brotman, Harris F.  
REGISTRATION NUMBER: 35,461  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 654-2428  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1581 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-09-320-774-6

Alignment Scores:  
Pred. No.: 0.161 Length: 1581  
Score: 90.50 Matches: 47  
Percent Similarity: 39.88% Conservative: 22  
Best Local Similarity: 27.17% Mismatches: 41  
Query Match: 11.65% Indels: 63  
DB: 4 Gaps: 8

US-09-847-539A-6 (1-159) x US-09-320-774-6 (1-1581)

QY 35 ArgAlaIleAspGluLeuLysGlnAla----- 44  
DB 925 CGCGTGTGGACAGCTGTGGAGAGCGCGCGCGCTGTGCAGCCAAATCGAGCACC 984  
QY 45 IleGluAspLysGluAlaThrThra----- 53  
DB 985 CTGGCGGACAAAGCCCGCCAGCAGCGCCAGCGTGCAGCGCTTGGAGCGGACGCC 1044  
QY 54 -----IleGluAla----- 56  
DB 1045 AAGATGAGCGCGAGCTGCAGAGGAGAGCTGAGCGCGTGGCGGCGCCACGTTT 1104  
QY 57 -----AlaSerSerAspAlaLeuGluAlaLeuAlaAsp 67  
DB 1105 GTGCTGCGCGAGCTCAAGCCCGCGCGCGCGCGCGCGCTGCGCGAGCGCGCGCG 1164  
QY 68 GlnThrAspAlaLeuGlnSerGluAlaAlaValValys----- 81  
DB 1165 GAGCTGACGCGG---CAGCGCGAGAGAGCGCCCAACGCCAAGTGAGAGCGGACAGCGC 1221  
QY 82 -----AlaAspAsnAlaIleSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAla 99  
DB 1222 GAGCGCGCGAGAGAGCGCGCGCGCGCGCG---GAGCGCGCGCGGAGAGAGCAGAGCGC 1278  
QY 100 LeuGlnSerGluGluAlaGlnValValGlnSerAspAsnAlaIleSerAspAlaTrpGlu 119  
DB 1279 CTGCTGAGAGGTGGCGCC-----ACGCGCGCGCGGAGAGAGAGAGAG 1323  
QY 120 LysAlaIleThrProIleAlaLeuAspValLysThrLysAspThrLysProVal--- 138  
DB 1324 CGCGCGAGAGAGCGCGCGCGCGCGCGCGCGCTGCGCGAGTGCGCGTGTGAC 1383  
QY 139 ValLysLysGluGluArgGlnAsnValAsnThrLeuPro 151  
DB 1384 GTGAGGCTGAGTGCCCAAGCGGTGAGAGCTGTGCC 1422

RESULT 23  
US-08-106-981-1  
Sequence 1, Application US/08106981  
Patent No. 573149  
GENERAL INFORMATION:  
APPLICANT: SARHAN, Fathey  
APPLICANT: SARHAN, Mario  
APPLICANT: LALIBERT, Jean-Fran ois  
TITLE OF INVENTION: DNA MOLECULES ENCODING FREEZING  
TITLE OF INVENTION: TOLERANCE PROTEINS IN GRAMINAE  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:

ADDRESSEE: NIXON & VANDERHIVE P.C.  
 STREET: 1100 NORTH GLEBE ROAD  
 CITY: ARLINGTON  
 STATE: VIRGINIA  
 COUNTRY: U.S.A.  
 ZIP: 22201-4714  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/106,981  
 FILING DATE: 16-AUG-1993  
 CLASSIFICATION: 530  
 ATTORNEY/AGENT INFORMATION:  
 NAME: WILSON, MARY J.  
 REGISTRATION NUMBER: 32,955  
 REFERENCE/DOCKET NUMBER: 163-23  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (703) 816-4000  
 TELEFAX: (703) 816-4100  
 TELEX: 200797 NIXN UR  
 INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 943  
 TYPE: Nucleic acid  
 STRANDEDNESS: Double  
 TOPOLOGY: Linear  
 MOLECULE TYPE: cDNA to mRNA  
 ORIGINAL SOURCE: Trillium aestivum L.  
 US-08-106-981-1

Alignment Scores:  
 Pred. No.: 0.0936 Length: 943  
 Score: 90.00 Matches: 36  
 Percent Similarity: 43.97% Conservative: 15  
 Best Local Similarity: 31.03% Mismatches: 45  
 Query Match: 11.58% Indels: 20  
 DB: 1 Gaps: 5

US-09-847-539A-6 (1-159) x US-08-106-981-1 (1-943)

QY 7 GlnProArgIleLeuProAsnGlyThrLeuThrAsnLeuLeuGlyAsnAlaProGlu 26  
 DB 433 CGCCCGCGACCGTG--AGCGGTGCGCTC-----TGCTGCAAGAAAGCTCCAGG 483  
 QY 27 LysLeuAlaLeuArgAsnGlu-----GluArgAlaIle----- 37  
 DB 484 CGCACTCCGCGCGTACACAGCAGCTGCGATGCCACCGACAGCCCTGCAGCGCGTGAAG 543  
 QY 38 -----AspGluLeuLysLysGlnAlaIleGluAspLysGlnAlaThr----- 52  
 DB 544 GCGGTGCGCGACGACGTTGAAGAAAGCGCGTGGCGAGCGTGGCGAGCCGCTCGGGCAAC 603  
 QY 53 -----AlaIleGluAlaIleAsnSerAspAlaLeuGluAlaLeuAlaAspGlnThr 69  
 DB 604 ACCGAAAGCGCGCGAGAGCGCGCAAGCGCGCGAGGTGAGCGCAAGCGCCAG 663  
 QY 70 AspAlaLeuGlnSerGlnGluAlaIleValLysAlaAspAsnAlaIleSerAspAla 89  
 DB 664 GACTTCGCGCAGCAGCAAGCAAGCGCGAGGAGGAGCGCTGGCGAGCGCCCAAGAGCGCC 723  
 QY 90 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGlnGluAla 105  
 DB 724 GCACAGGCGATCAGGACAAAGTCGCCGCCGCCCAAAAAGAGACCT 771

# RESULT 24

US-08-072-070-1  
 Sequence 1, Application US/08072070  
 Patent No. 5476929  
 GENERAL INFORMATION:  
 APPLICANT: Briles, David E

APPLICANT: Volter, Janet L  
 APPLICANT: McDaniel, Larry S  
 TITLE OF INVENTION: STRUCTURAL GENE OF PNEUMOCOCCAL  
 TITLE OF INVENTION: PROTEIN  
 NUMBER OF SEQUENCES: 6  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Shoemaker and Mattare, Ltd  
 STREET: Suite 1203, 2001 Jefferson Davis Highway  
 CITY: Arlington  
 STATE: Virginia  
 COUNTRY: U.S.A.  
 ZIP: 22202-0286  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/072,070  
 FILING DATE: 19930603  
 CLASSIFICATION: 424  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US/07/835,698  
 FILING DATE: 12-FEB-1992  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US/07/656,773  
 FILING DATE: 15-FEB-1991  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (703) 415-0810  
 TELEFAX: (703) 521-0378  
 TELEX: LUKPAT WASHINGTON  
 INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 2085 base pairs  
 TYPE: NUCLEIC ACID  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 HYPOTHETICAL: NO  
 ANTI-SENSE: NO  
 ORIGINAL SOURCE:  
 ORGANISM: Streptococcus pneumoniae  
 STRAIN: Rxi1  
 IMMEDIATE SOURCE:  
 CLONE: JY2008  
 FEATURE:  
 NAME/KEY: Intron  
 LOCATION: 1..2085  
 NAME/KEY: CDS  
 LOCATION: join(127..1983, 1987..1992, 1996..2007, 2011  
 LOCATION: ..2025, 2029..2031, 2035..2085)  
 US-08-072-070-1

Alignment Scores:  
 Pred. No.: 0.263 Length: 2085  
 Score: 90.00 Matches: 37  
 Percent Similarity: 39.22% Conservative: 23  
 Best Local Similarity: 24.18% Mismatches: 61  
 Query Match: 11.58% Indels: 32  
 DB: 1 Gaps: 4

US-09-847-539A-6 (1-159) x US-08-072-070-1 (1-2085)

QY 24 AlaProGluLysLeuAlaLeuArgAsnGluGluArgAla----IleAspGluLeuLysLys 42  
 DB 634 GCACAGAACTTACTTAAATACTAGAAAGCTTAAAGCAAAATAGAAAGGCTGAGAAA 693  
 QY 43 GlnAlaIleGluAspLysGlnAlaThrThrAlaIleGluAlaIleSerSerAspAlaLeu 62  
 DB 694 AAAGCTACTGAAGCCCAACAAATAAGTGGATGCTGAAGAAAGTGGCTCCTCAACTAAATC 753  
 QY 63 GluAlaIleuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaIleVal----- 79





```

SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/214,164
FILING DATE: 17-MAR-1994
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/656,773
FILING DATE: 15-FEB-1991
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W.
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 6102-137
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 521-0813
TELEX: LUKPAT WASHINGTON
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2085 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: Intron
LOCATION: 1..1983
FEATURE:
NAME/KEY: CDS
LOCATION: 127..1983
US-08-214-164-1

Alignment Scores:
Pred. No.: 0.263 Length: 2085
Score: 90.00 Matches: 37
Percent Similarity: 39.22% Conservative: 23
Best Local Similarity: 24.18% Mismatches: 61
Query Match: 11.58% Indels: 32
Gaps: 4

US-09-847-539a-6 (1-159) x US-08-214-164-1 (1-2085)
QY 24 AAlaProGluLysLeuAlaLeuAArgGluGluArgAla----IleAspGluLeuLysLys 42
Db 634 GCACCAAGAACTACTAAAAAAGTGAAGAGCTAAAGCAAAATTAGAGAGCGCTGAGAAA 693
QY 43 GlnAlaIleGluAspLysGluAlaThrThrAlaIleGluAlaIleSerSerAspAlaLeu 62
Db 694 AAAGCTACTGAGCAACCAAAAGTGAAGTCTGAAGAGAGTGGCTCTCAAGCTAAATATC 753
QY 63 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaIleVal----- 79
Db 754 GCTGAATTGCAAAATCAAGTTCATAGACTAGAACACAGCTCAAGAGATTGAGTCT 813
QY 80 -----ValLysAla 82
Db 814 CAATCAGAGATTATGCTAAAGAGGTTCCGCTCCTCTTCAATCTAAATTGAGATGCC 873
QY 83 AspAsnAlaIleSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSer 102
Db 874 AAAAAGCTAAACTATATAAAAGTGAAGAGTTAGTGATGATGATGATGATGATGATGATG 933
QY 103 GluGluAlaGluAlaValGlnSerAspAsnAlaIle----- 114
Db 934 GAATTCGCAAAACTTGAAGATCAACTTAAAGCTCTGAGAGAAACAAATATGAGAGAC 993
QY 115 -----SerAspAlaIlePepGluLysAlaIleThrProIleAlaLeuAspValLysLysThr 132
Db 994 TACTTTAAAGAAAGCTTTAGAGAAAGCTATTGCTGCTAAAAAGCTGCAATTAGAAAAAACT 1053
QY 133 Lys---AspThrLysProValLysLysLysLysLysLys 144
Db 1054 GAAAGCTAGCTTAAAGAAAGCAAGTTAATGAGCGAGAAAAA 1092

```

```

RESULT 27
US-08-469-434-1
Sequence 1, Application US/08469434
Patent No. 5753463
GENERAL INFORMATION:
APPLICANT: Billies, David E
APPLICANT: Youther, Janet L
APPLICANT: McDaniel, Larry S
TITLE OF INVENTION: STRUCTURAL GENE OF PNEUMOCOCCAL
TITLE OF INVENTION: PROTEIN
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Shoemaker and Mattare, Ltd
STREET: Suite 1203, 2001 Jefferson Davis Highway
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/469,434
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/072,065
FILING DATE: 03 JUNE 1993
APPLICATION NUMBER: US/07/835,698
FILING DATE: 12-FEB-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/656,773
FILING DATE: 15-FEB-1991
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 521-0378
TELEX: LUKPAT WASHINGTON
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2085 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Streptococcus pneumoniae
STRAIN: Rxi
IMMEDIATE SOURCE:
CLONE: JY2008
FEATURE:
NAME/KEY: Intron
LOCATION: 1..2085
FEATURE:
NAME/KEY: CDS
LOCATION: join(127..1983, 1987..1992, 1996..2007, 2011
LOCATION: ...2025, 2029..2031, 2035..2085)
US-08-469-434-1

Alignment Scores:
Pred. No.: 0.263 Length: 2085
Score: 90.00 Matches: 37
Percent Similarity: 39.22% Conservative: 23
Best Local Similarity: 24.18% Mismatches: 61
Query Match: 11.58% Indels: 32
Gaps: 4

US-09-847-539a-6 (1-159) x US-08-469-434-1 (1-2085)

```

```

QY 24 AlaprogululsleuAlaleuArgnsclugluArgala---lleaspluleuylslys 42
      |||||
Db 634 GCACGAGAACTTACTTAAAAAAGCTAGAGAGCTAAAGCAAAATTAGAGAGGCTAGAAA 693
QY 43 GlnAlaIleGluAspLysGluAlaThrAlaIleGluAlaIleSerSerAspAlaLeu 62
      |||||
Db 694 AAAGCTACTGAGCAAAAGCAAAAGTGAGTGTGAGAGAGCGCTCTCAAGCTAAAATC 753
QY 63 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaIleVal----- 79
      |||
Db 754 GCTGAATTGAAAAATCAAGTTCATAGACTAGAACAGAGCTCAAGAGATTGATGCT 813
QY 80 -----ValysAla 82
      |||
Db 814 GAATCAGAAAGATTATGCTAAGAAAGTTCCGTCCTCTTCATCTAAATTTGATGCTC 873
QY 83 AspaAlaIleAlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSer 102
      |||
Db 874 AAAAAAGCTAAAGCTATCAAAAGCTTGAAGAGTTAACTGATTAAGATTGATGAGCT 933
QY 103 GluGluAlaGluValValGlnSerAspAlaIle----- 114
      |||
Db 934 GAATTTGCAAAAGCTTGAAGATCAACTTAAAGCTGCTGAAGAAACAAATTAATGTAGAAGC 993
QY 115 -----SerAspAlaTrpGluLysAlaIleAlaThrProIleAlaLeuAspValLysThr 132
      |||||
Db 994 TACTTTAAGAAAGTTTATAGAGAAAGCTATGCTGCTTAAAAAGCTGATTAAGAAAACT 1053
QY 133 Lys---AspThrLysProValValLysLysGluGluArg 144
      |||
Db 1054 GAAGCTGACCTTAAGAAAGAGCTTAATGAGCCAGAAAAA 1092

```

## RESULT 28

US-08-214-222-1

```

; Sequence 1, Application US/08214222
; Patent No. 5804193
; GENERAL INFORMATION:
; APPLICANT: Briles, David E
; TITLE OF INVENTION: STRUCTURAL GENE OF PNEUMOCOCCAL
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Maltare, Ltd
; STREET: Suite 1203, 2001 Jefferson Davis Highway
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/214,222
; FILING DATE: 17-MAR-1994
; CLASSIFICATION: 530
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 07/835,698
; FILING DATE: 12-FEB-1992
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 321-0378
; TELEX: LUKPAT WASHINGTON
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2085 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO

```

```

; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pneumoniae
; STRAIN: Rx1
; IMMEDIATE SOURCE:
; CLONE: JY2008
; FEATURE:
; NAME/KEY: intron
; LOCATION: 1..2085
; NAME/KEY: CDS
; LOCATION: join(127..1983, 1987..1992, 1996..2007,
; LOCATION: 2011
; LOCATION: ..2025, 2029..2031, 2035..2085)
; US-08-214-222-1

```

## Alignment Scores:

Pred. No.:	0.263	Length:	2085
Score:	90.00	Matches:	37
Percent Similarity:	39.22%	Conservative:	23
Best Local Similarity:	24.18%	Mismatches:	61
Query Match:	11.58%	Indels:	32
DB:	1	Gaps:	4

US-09-847-539a-6 (1-159) x US-08-214-222-1 (1-2085)

```

QY 24 AlaprogululsleuAlaleuArgnsclugluArgala---lleaspluleuylslys 42
      |||||
Db 634 GCACGAGAACTTACTTAAAAAAGCTAGAGAGCTAAAGCAAAATTAGAGAGGCTAGAAA 693
QY 43 GlnAlaIleGluAspLysGluAlaThrAlaIleGluAlaIleSerSerAspAlaLeu 62
      |||||
Db 694 AAAGCTACTGAGCAAAAGCAAAAGTGAGTGTGAGAGAGCGCTCTCAAGCTAAAATC 753
QY 63 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaIleVal----- 79
      |||
Db 754 GCTGAATTGCAAAAGCTTGAAGATCAACTTAAAGCTGCTGAAGAAACAAATTAATGTAGAAGC 993
QY 80 -----ValysAla 82
      |||
Db 814 GAATCAGAAAGATTATGCTAAGAAAGTTCCGTCCTCTTCATCTAAATTTGATGCTC 873
QY 83 AspaAlaIleAlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSer 102
      |||
Db 874 AAAAAAGCTAAAGCTATCAAAAGCTTGAAGAGTTAACTGATTAAGATTGATGAGCT 933
QY 103 GluGluAlaGluValValGlnSerAspAlaIle----- 114
      |||
Db 934 GAATTTGCAAAAGCTTGAAGATCAACTTAAAGCTGCTGAAGAAACAAATTAATGTAGAAGC 993
QY 115 -----SerAspAlaTrpGluLysAlaIleAlaThrProIleAlaLeuAspValLysThr 132
      |||||
Db 994 TACTTTAAGAAAGTTTATAGAGAAAGCTATGCTGCTTAAAAAGCTGATTAAGAAAACT 1053
QY 133 Lys---AspThrLysProValValLysLysGluGluArg 144
      |||
Db 1054 GAAGCTGACCTTAAGAAAGAGCTTAATGAGCCAGAAAAA 1092

```

## RESULT 29

US-08-467-852A-1

```

; Sequence 1, Application US/08467852A
; Patent No. 5856170
; GENERAL INFORMATION:
; APPLICANT: BRILES, David E.
; APPLICANT: YOTTER, Janet L.
; APPLICANT: MCDANIEL, Larry S.
; TITLE OF INVENTION: STRUCTURAL GENE OF PNEUMOCOCCAL PROTEIN
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FROMMER LAWRENCE & HAUG LLP
; STREET: 745 Fifth Avenue
; CITY: New York
; STATE: NY

```

Oy	115	-----SerAspAlatPrgLysValIalaAthrProillealaleuapvalylsYstR	132
		:::	:::
Dd	994	TACTTTAAAGACAGCTTTCAGGAAAACATTCGTCCGTCAAAAAAGCCTGAATTACAAAAACT	105
		:::	:::
Oy	133	Lys---AspThrLysProvalValLysLysGluGuaArg	144
		:::	:::
Dd	1054	GAACTGCACCTTAAGAAGACGTTAATGACCAGAAAAA	1092
		:::	:::
<hr/>			
	RESULT 30		
	US-08-468-718-1		
	: Sequence 1, Application US/08468718		
	: Patent No. 5871943		
	: GENERAL INFORMATION:		
	: APPLICANT: Briles, David E		
	: APPLICANT: Yother, Janet L		
	: APPLICANT: McDaniel, Larry S		
	: TITLE OF INVENTION: STRUCTURAL GENE OF PNEUMOCOCCAL		
	: NUMBER OF SEQUENCES: 6		
	: CORRESPONDENCE ADDRESS:		
	: ADDRESSEE: Shoemaker and Mattare, Ltd		
	: STREET: Suite 1203, 2001 Jefferson Davis Highway		
	: CITY: Arlington		
	: STATE: Virginia		
	: COUNTRY: U.S.A.		
	: ZIP: 22202-0286		
	: COMPUTER READABLE FORM:		
	: MEDIUM TYPE: Floppy disk		
	: COMPUTER: IBM PC compatible		
	: OPERATING SYSTEM: PC-DOS/MS-DOS		
	: SOFTWARE: PatentIn Release #1.0, Version #1.25		
	: CURRENT APPLICATION DATA:		
	: APPLICATION NUMBER: US/08/468,718		
	: FILING DATE:		
	: CLASSIFICATION: 435		
	: PRIOR APPLICATION DATA:		
	: APPLICATION NUMBER: US/08/072,068		
	: FILING DATE: 03 JUNE 1993		
	: APPLICATION NUMBER: US/07/835,698		
	: FILING DATE: 12-FEB-1992		
	: PRIOR APPLICATION DATA:		
	: APPLICATION NUMBER: US/07/656,773		
	: FILING DATE: 15-FEB-1991		
	: TELECOMMUNICATION INFORMATION:		
	: TELEPHONE: (703) 415-0810		
	: TELEFAX: (703) 521-0378		
	: TELEX: LUKPAT WASHINGTON		
	: INFORMATION FOR SEQ ID NO: 1:		
	: SEQUENCE CHARACTERISTICS:		
	: LENGTH: 2085 base pairs		
	: TYPE: nucleic acid		
	: STRANDEDNESS: double		
	: TOPOLOGY: linear		
	: MOLECULE TYPE: protein		
	: HYPOTHETICAL: NO		
	: ANTI-SENSE: NO		
	: ORIGINAL SOURCE:		
	: ORGANISM: Streptococcus pneumoniae		
	: STRAIN: Rxi		
	: IMMEDIATE SOURCE:		
	: CLONE: jY2008		
	: FEATURE:		
	: NAME/KEY: Intron		
	: LOCATION: 1..2085		
	: FEATURE:		
	: NAME/KEY: CDS		
	: LOCATION: Join(127..1983, 1987..1992, 1996..2007, 2011		
	: LOCATION: ..2025, 2029..2031, 2035..2085)		
	: US-08-468-718-1		
<hr/>			
Alignment Scores:	0.263	Length:	2085
red. No.:			

```

Score: 90.00 Matches: 37
Percent Similarity: 39.22% Conservative: 23
Best Local Similarity: 24.18% Mismatches: 61
Query Match: 11.58% Indels: 32
DB: 2 Gaps: 4

US-09-847-539a-6 (1-159) x US-08-468-718-1 (1-2085)

QY 24 AlaProgluylsleuAlaLeuArgAsnGluAlaArgAla---IleAspGluLeuLysLys 42
DB 634 GCACCGAAGACTTACTTAAAAAAGCTAGAGAGCTAAAGCAAAATTAGAGAGGCTGAGAAA 693
QY 43 GluAlaIleGluAspLysGluAlaThrThrAlaIleGluAlaIleSerSerAspAlaLeu 62
DB 694 AAAGCTGCTAGAGCAACCAAAAGTGATGCTGAGAGAGTGGCTCCTCAAGCTAAATC 753
QY 63 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaVal----- 79
DB 754 GCTGAATTGGAAATCAAGTTTCATAGACTAGAACAGACGTCANAGAGATTGATGAGTCT 813
QY 80 -----ValLysAla 82
DB 814 GAATCAGAGATTATGCTAAAGAGGTTTCGCTCCTCTTCAATCTAAATTGATGAGCC 873
QY 83 AspAsnAlaIleSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSer 102
DB 874 AAAAAGCTAAACTATCAAAAGCTTGAAGAGTTAAGTATGATGATGATGAGGCT 933
QY 103 GluGluAlaGluValValGlnSerAspAlaAla----- 114
DB 934 GAATTTGCAAAAGCTTGAAGATCAACTAAAGCTGCTGAGAGAAACATATGTAAGAGAC 993
QY 115 -----SerAspAlaTrpGluLysAlaAlaThrProIleAlaLeuAspValLysLysThr 132
DB 994 TACTTTAAAGAGCTTGAAGAGAAAGCTATGCTGCTAAAGAGCTGAATTGAGAAAAACT 1053
QY 133 Lys---AspThrLysProValValLysGluGluArg 144
DB 1054 GAAGCTGACCTTAAGAAAGCAGTTAATGAGCCAGAAAAA 1092

RESULT 31
US-08-246-636-1
; Sequence 1, Application US/08246636
; Patent No. 5965141
; GENERAL INFORMATION:
; APPLICANT: Biles, David E
; APPLICANT: Yocheer, Janet L
; APPLICANT: McDaniel, Larry S
; APPLICANT: Wu, Hong-Yin
; TITLE OF INVENTION: EPITOPIC REGIONS OF PNEUMOCOCCAL SURFACE
; TITLE OF INVENTION: PROTEIN A
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Matlare, Ltd
; STREET: Suite 1203, 2001 Jefferson Davis Highway
; City: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/246,636
; FILING DATE: 20-MAY-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/656,773
; FILING DATE: 15-FEB-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/835,698

```

```

; FILING DATE: 12-FEB-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/048,896
; FILING DATE: 20-APR-1993
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; TELEX: LUKPAT WASHINGTON
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2085 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pneumoniae
; STRAIN: Rxi
; IMMEDIATE SOURCE:
; CLONE: JY4313
; FEATURE:
; NAME/KEY: Intron
; LOCATION: 1..2085
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join(127..1984)
; US-08-246-636-1

Alignment Scores:
Score: 0.263 Length: 2085
Percent Similarity: 39.22% Matches: 37
Best Local Similarity: 24.18% Conservative: 23
Query Match: 11.58% Mismatches: 61
DB: 2 Indels: 32
Gaps: 4

US-09-847-539a-6 (1-159) x US-08-246-636-1 (1-2085)

QY 24 AlaProgluylsleuAlaLeuArgAsnGluAlaArgAla---IleAspGluLeuLysLys 42
DB 634 GCACCGAAGACTTACTTAAAAAAGCTAGAGAGCTAAAGCAAAATTAGAGAGGCTGAGAAA 693
QY 43 GluAlaIleGluAspLysGluAlaThrThrAlaIleGluAlaIleSerSerAspAlaLeu 62
DB 694 AAAGCTGCTAGAGCAACCAAAAGTGATGCTGAGAGAGTGGCTCCTCAAGCTAAATC 753
QY 63 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaVal----- 79
DB 754 GCTGAATTGGAAATCAAGTTTCATAGACTAGAACAGACGTCANAGAGATTGATGAGTCT 813
QY 80 -----ValLysAla 82
DB 814 GAATCAGAGATTATGCTAAAGAGGTTTCGCTCCTCTTCAATCTAAATTGATGAGCC 873
QY 83 AspAsnAlaIleSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSer 102
DB 874 AAAAAGCTAAACTATCAAAAGCTTGAAGAGTTAAGTATGATGATGATGAGGCT 933
QY 103 GluGluAlaGluValValGlnSerAspAlaAla----- 114
DB 934 GAATTTGCAAAAGCTTGAAGATCAACTAAAGCTGCTGAGAGAAACATATGTAAGAGAC 993
QY 115 -----SerAspAlaTrpGluLysAlaAlaThrProIleAlaLeuAspValLysLysThr 132
DB 994 TACTTTAAAGAGCTTGAAGAGAAAGCTATGCTGCTAAAGAGCTGAATTGAGAAAAACT 1053
QY 133 Lys---AspThrLysProValValLysGluGluArg 144
DB 1054 GAAGCTGACCTTAAGAAAGCAGTTAATGAGCCAGAAAAA 1092

RESULT 32

```



```

: ANTI-SENSE: NO
: ORIGINAL SOURCE:
: ORGANISM: Streptococcus pneumoniae
: STRAIN: Rxi
: IMMEDIATE SOURCE:
: CLONE: JY4313
: FEATURE:
: NAME/KEY: Intron
: LOCATION: 1..2085
: FEATURE:
: NAME/KEY: CDS
: LOCATION: join(127..1984)
: US-08-319-795-1

Alignment Scores:
Pred. No.: 0.263 Length: 2085
Score: 90.00 Matches: 37
Percent Similarity: 39.22% Conservative: 23
Best Local Similarity: 24.18% Mismatches: 61
Query Match: 11.58% Indels: 32
DB: 2 Gaps: 4

US-09-847-539a-6 (1-159) x US-08-319-795-1 (1-2085)
QY 24 AlaprogilulsleuAlaleuArgasnGlucIuArgala---lleaspGluleuLyslys 42
DB 634 GCACCAAGACTTACTAAAAAACTAGAAAGAGCTAAAGCAAAATTAGAGAGCGTGAGAAA 693
QY .43 GluAlaIleGlualAspLysGluAlaThrThrAlaIleGlualAlaSerSerAspAlaLeu 62
DB 694 AAAGCTACTGAAAGCCAAACAAAGGTGATGAGAGAGTCGCTCCCAAGCTAAATC 753
QY 63 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaVal----- 79
DB 754 GCTGAATTGGAATTCAGATTCTATGACAGACAGCTCAAGAGATGATGACTGT 813
QY 80 -----VallysAla 82
DB 814 GAATCAGAGATTTGCTAAAGAGTTCCGTCCTCTTCATCTCAATTTGGATGCC 873
QY 83 AspAsnAlaAlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSer 102
DB 874 AAAAAGCTAAACTATCAAACTGAAAGAGTTAGTATGATGATGATGATGATGATGATG 933
QY 103 GluGluAlaGluValValGlnSerAspAsnAlaAla----- 114
DB 934 GAAATTCGCAAACTGGAAGATCAACTAAAGCTGCTGAAGAAACAAATTAATGTAAGAC 993
QY 115 -----SerAspAlaTpgIuLysAlaAlaAlaThrProIleAlaLeuAspValLysLysThr 132
DB 994 TACTTTAAAGAGTTTAAAGAAACCTATTTCTGCTAAAAAAGCTGAATTTGAAAAAAGCT 1053
QY 133 Lys---AspThrLysProValValLysLysGluLysArg 144
DB 1054 GAAGCTGACCTTAAAGAAAGCAGTTAATGAGCCAGAAAAA 1092

RESULT 34
: US-08-468-985-1
: Sequence 1, Application US/08468985
: Patent No. 5997882
: GENERAL INFORMATION:
: APPLICANT: Briles, David E.
: APPLICANT: Yother, Janet L.
: APPLICANT: McDaniel, Larry S
: TITLE OF INVENTION: Epitopic Regions of Pneumococcal Surface
: TITLE OF INVENTION: Protein A
: NUMBER OF SEQUENCES: 20
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Sheomaker and Mattare, Ltd.
: STREET: 1203 Crystal Plaza Bldg. 1, 2001 Jefferson
: STREET: Davis Highway
: CITY: Arlington
: STATE: Virginia

```

```

: COUNTRY: U S A.
: ZIP: 22202-0286
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/468,985
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/319,795
: FILING DATE:
: APPLICATION NUMBER: US 08/246,636
: FILING DATE: 20-MAY-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/048,896
: FILING DATE: 20-APR-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/835,698
: FILING DATE: 12-FEB-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/656,773
: FILING DATE: 15-FEB-1991
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703) 415-0810
: TELEFAX: (703) 415-0813
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2085 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: HYPOTHEICAL: NO
: ANTI-SENSE: NO
: ORIGINAL SOURCE:
: ORGANISM: Streptococcus pneumoniae
: STRAIN: Rxi
: IMMEDIATE SOURCE:
: CLONE: JY4313
: FEATURE:
: NAME/KEY: Intron
: LOCATION: 1..2085
: FEATURE:
: NAME/KEY: CDS
: LOCATION: join(127..1984)
: US-08-468-985-1

Alignment Scores:
Pred. No.: 0.263 Length: 2085
Score: 90.00 Matches: 37
Percent Similarity: 39.22% Conservative: 23
Best Local Similarity: 24.18% Mismatches: 61
Query Match: 11.58% Indels: 32
DB: 2 Gaps: 4

US-09-847-539a-6 (1-159) x US-08-468-985-1 (1-2085)
QY 24 AlaprogilulsleuAlaleuArgasnGlucIuArgala---lleaspGluleuLyslys 42
DB 634 GCACCAAGACTTACTAAAAAACTAGAAAGAGCTAAAGCAAAATTAGAGAGCGTGAGAAA 693
QY .43 GluAlaIleGlualAspLysGluAlaThrThrAlaIleGlualAlaSerSerAspAlaLeu 62
DB 694 AAAGCTACTGAAAGCCAAACAAAGGTGATGAGAGAGTCGCTCCCAAGCTAAATC 753
QY 63 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaVal----- 79
DB 754 GCTGAATTGGAATTCAGATTCTATGACAGACAGCTCAAGAGATGATGACTGT 813
QY 80 -----VallysAla 82

```

```

Db      814 GAATCAAGAAATTATGCTTAAAGAAGTTTCCGTGCTCCTTCCTTAATCATTAATGATGCC      87
Oy      03 Aspsmnlalabseraspalaicenglualaleualaaspglnthrspalalainser      10
           |||          |||||         |:::||||:    || |||   ::
Db      874 AAAAAGCTAAACTATCAAAACCTTGAAAGAGTTAAAGATGATGAATGAGTAGAACGCT      93
Oy      115 -----setaspalatrpclulsalsaalathrPcollalaleuaspaVlLysLysThr      13
           |||          |::          |||||         |:::||||:    |||||
Db      934 GAAATGCAAAACCTGACATCACTTAAAGCTGCTGCAAGAAAAACATAATGTAGAAGAC      99
Oy      994 TACTTTAAAGAAAGTTTAAAGAAACCTATTGGCTGTCTAAAAAAGCTGAATTAGAAAAAACT      10
Oy      133 Lys---AspthlysprprovalValLylsLysglGuatrg      144
           ::| ||| ||| ||| ||| ::| ||||:
Db      1054 GAAGCTGACCTTAAGAAACAGTTAATGAGCACGAAAAA      1092

RESULT 35
US-08-312-949-1
; Sequence 1, Application US/08312949
; Patent No. 6027734
GENERAL INFORMATION:
APPLICANT: Biles, David E.
APPLICANT: Wu, Hong-Yin
TITLE OF INVENTION: MUCOSAL ADMINISTRATION OF
TITLE OF INVENTION: PNEUMOCOCCAL ANTIGENS
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSER: Curtis, Morris & Safford, P.C.
STREET: 530 Fifth Avenue
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/312,949
FILING DATE: 30-SEP-1994
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Frommer, William S.
REGISTRATION NUMBER: 25,506
REFERENCE/DOCKET NUMBER: 454312-2049
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 840-3333
TELEFAX: (212) 840-0712
INFORMATION FOR SEQ ID NO.: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2086 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-312-949-1

Alignment Scores:
Pred. No.:      0.263      Length:      2086
Score:          90.00     Matches:       37
Percent Similarity: 39.22%   Conservative: 23
Best Local Similarity: 24.18% Mismatches:   61
Query Match:      11.56%    Indels:        32
DB:               3        Gaps:           4

US-09-847-539A-6 (1-159) x US-08-312-949-1 (1-2086)
Oy      24 AlaPrGluLysLeuAlaLeuArGasnGlucLUAArgLa---LIlaSpGIuLeuLysLys      42
           |||||
           |||  ::|||  |||:|||||  :|||:|||||  ::|||

```

```

Db      634 GCACGACGAAGTCTACTAAAAAACTCGAAGAAGCCTAAACCAAAATTTAGAAAGGGCTGAGAAA   693
Qy      43 GlnAlalleguaaplysgluAlaIthrThralIllegluAlaIasSeraspAlaLeu    62
       ::::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      694 AAACCTTGTAAGGCCAACCAAAAGTGGATGTGAAGAAGTCGCTCTCAAGCTAAATAATTC   753
Qy      63 GluAlaIeuaIaaspGlnThrAspAlaLeugInserGluAlaIaIaVal-----79
       || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      754 GCTGAATTGSAANAATCAAGTTCATAGACTGAAACAAGAGCTCAAAGAGATTGAGACTT    813
Qy      80 -----ValysAla 82
       || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      814 GAATCAGAAGATTATGCTAAAGAAGCTTCGTCCTCTCAATCTAAATTGGATGCC     873
Qy      83 AspsnAlaIasSeraspAlaLeuGlAlaIeuaIaaspGlnThrAspAlaLeuGInser    102
       || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      874 AAAAAGCTTAACATCTCAAAACTGGAAGAGATTAAAGTATTAAGTTGATGAGACGCT    933
Qy      103 GluGluaIagluValglInSeraspAlaIaIa-----114
       || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      934 GAAATTGCAGAAACTGGAAGATCAACTTAAGCTCTGAAGAAACATTAATGTAGAGAC    993
Qy      115 -----SeraspAlaIatTPGluLyAlaIaIaIThrIleAlaIeuaSpvLysLyThr    132
       ::::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      994 TACTTTAAAGAGGTTTAGAGAAAACCTATTGCTGCTAAAAAGCTGAATTAGAAAAAACT    1053
Qy      133 Lys---AspThrLysProvalValysLysGluLyuar 144
       ::::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1054 GAAGCTGACCTTAGAAAAGCAGTTAAATAGAGCCAGAAAAA 1092

RESULT 36
US-08-446-201-2
; Sequence 2, Application US/08446201B
; Patent No. 6042838
; GENERAL INFORMATION:
; APPLICANT: BRILES, David E.
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS FOR MUCOSAL ADMINISTRATION OF
; FILE REFERENCE: 454312-2018
; CURRENT APPLICATION NUMBER: US/08/446,201B
; EARLIER FILING DATE: 1995-05-19
; EARLIER APPLICATION NUMBER: 08/312,949
; EARLIER FILING DATE: 1994-09-30
; EARLIER APPLICATION NUMBER: 08/246,636
; EARLIER FILING DATE: 1994-05-20
; EARLIER APPLICATION NUMBER: 08/048,896
; EARLIER FILING DATE: 1993-04-20
; EARLIER APPLICATION NUMBER: 07/835,698
; EARLIER FILING DATE: 1992-02-12
; EARLIER APPLICATION NUMBER: 07/656,773
; EARLIER FILING DATE: 1991-02-15
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Patentin Ver. 2.0
SEO ID NO 2
LENGTH: 2086
TYPE: DNA
ORGANISM: Streptococcus pneumoniae
US-08-446-201-2

Alignment Scores:
Pred.: 0.263 Length: 2086
Score: 90.00 Matches: 37
Percent Similarity: 39.22% Conservative: 23
Best Local Similarity: 24.18% Mismatches: 61
Query Match: 11.58% Indels: 32
DB: 3 Gaps: 4

US-09-847-539A-6 (1-159) x US-08-446-201-2 (1-2086)
Qy      24 AlaProGluLysLeuAlaIeuaTrgAsrGluGluArgAla----IleaspGluLeuLys 42
       ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      634 GCACGACGAAGTCTACTAAAAAACTAGAAAGCCTAAACCAAAATTTAGAAAGGGCTGAGAAA   693

```



```
QY 43 GlnAlaIleGluAspLysGluAlaThrThrAlaIleGluAlaIleSerSerAspAlaLeu 62
Db 694 AAAAGCTACTAGACCAACAAAGAGTGAAGAGCTGCTCTCAAGCTAAATC 753
QY 63 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaVal----- 79
Db 754 GGTGAATTCGAATATCAAGATTAGACTAGAACAGAGCTCAAGAGATTGATGACTCT 813
QY 80 -----ValLysAla 82
Db 814 GAATCAAGAGATTATGCTAAAGAGTTTCCGTCCTCTCTCAATCAATTAATTGATGCC 873
QY 83 AspAsnAlaIleSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSer 102
Db 874 AAAAAGCTTAAGCTATCAAAAGAGTGAAGAGTGAAGATTAAGATTAAGATTAAGAGCT 933
QY 103 GluGluAlaGluAlaValGlnSerAspAsnAlaIle----- 114
Db 934 GAATTCGAAGAGCTGAAGCTTAAGCTGCTGAGAGAAACATTAATGCTGAAGAC 993
QY 115 -----SerAspAlaTrpGluLysAlaIleThrProIleAlaLeuAspValLysThr 132
Db 994 TACTTTAAAGAGGTTTGAAGAAAGCTATGCTGCTAAAGAGCTCAATTAAGAAAGCT 1053
QY 133 Lys---AspThrLysProValValLysLysGluGluArg 144
Db 1054 GAAGCTGACCTTAAGAAAGCAGTTAATGAGCCAGAAAAA 1092

RESULT 37
US-08-937-271-9
; Sequence 9, Application US/08937271
; Patent No. 6063386
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; TITLE OF INVENTION: RECOMBINANT MULTIVALENT M PROTEIN
; TITLE OF INVENTION: VACCINE
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/937,271
; FILING DATE: 15-SEP-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Roseman, Stephen J.
; REGISTRATION NUMBER: 43,058
; REFERENCE/DOCKET NUMBER: 481112, 405C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 918 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pyogenes
```

```
FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..918
; US-08-937-271-9

Alignment Scores:
Pred. No.: 0.134 Length: 918
Score: 88.50 Matches: 44
Percent Similarity: 37.02% Conservative: 23
Best Local Similarity: 24.31% Mismatches: 69
Query Match: 11.39% Indels: 45
DB: 3 Gaps: 5

US-09-847-539a-6 (1-159) x US-08-937-271-9 (1-918)
QY 19 AsnLeuGlnLysAlaIleProGluLysLeuAlaLeuArgAsnGluArgAlaIleAsp 38
Db 313 AACAAAGTTTCAGAAAGAGTGCAGAA--GGCTTCGCGGTGATTAGACGCATCAGCT 369
QY 39 GluLeuLysLysGluAlaIleGluAspLysGluAlaThrThrAlaIleGluAlaIleSer 58
Db 370 GAAGCTAAGAGCAATTAGAAAGCTGAACACAAACTTGAAGAACAAAGAGTTTCA 429
QY 59 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaIle 78
Db 430 GAACCAAGCTCCCAAGGCTTCGCCGCTGATTAGACGCATCAGCTGAAGCAAGCAAA 489
QY 79 ValValLysAlaAspAsnAlaIleSerAspAlaLeuGluAlaLeuAlaAspGlnThrAsp 98
Db 490 GTTGAAAAGCTTTAGAAAGAGCAAAACGCAAAATTAGCTGCTGTGAAAAAATTAAACAA 549
QY 99 AlaLeu-----GlnSerGluGluAlaGluValValGlnSerAsp 111
Db 550 GAGCTTGAAAGAAAGCAAAATTAAACAGAAAAAGAGCTGACCTCAAGCAAAACTT 609
QY 112 AsnAlaIleSerAspAlaTrpGluLysAlaIleThrProIleAlaLeuAspValLysThr 131
Db 610 GAACGACAGAAAGCAAGCTCAAGAAACAAATTAGCAAAACAGAGTGAAGAACTTGCAAAA 669
QY 132 -----ThrLysAspThrLysPro----- 137
Db 670 CTAAGAGCTGAAAAAGCATCAGACTCAACAAACCTGATACAAACGAAACAAAGCT 729
QY 138 -----Val 138
Db 730 GTTCAGGTAAAGTCAAGCACCAAGCAGTACAAACCAAAACAAAGCACCA 789
QY 139 ValLysLysGluGluArgGlnAsnValAsnThrLeuProThrThrGluGluSerAsn 158
Db 790 ATGAAGGAAGAACTAAGAGACAG-----TTACCATCAACAGGTGAAGAACACTAAC 837
QY 159 Pro 159
Db 838 CCA 840

RESULT 38
US-07-813-584A-1
; Sequence 1, Application US/07813584A
; Patent No. 5352588
; GENERAL INFORMATION:
; APPLICANT: Fischetti, Vincent A.
; APPLICANT: Bessen, Debra E.
; TITLE OF INVENTION: No. 5352588e1 Immunoglobulin A Binding Protein
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kiltie Murray
; STREET: 98 Cutler Mill Road
; CITY: Great Neck
; STATE: NY
; COUNTRY: USA
; ZIP: 11021
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```

```

: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/07/813,584A
: FILING DATE: 19911224
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: MURRAY, KITTLE
: REGISTRATION NUMBER: 30,246
: REFERENCE/DOCKET NUMBER: RU-100
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 516-482-1990
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1560 base pairs
: TYPE: NUCLEIC ACID
: STRANDEDNESS: single
: TOPOLOGY: unknown
: MOLECULE TYPE: DNA (genomic)
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 2..40
: NAME/KEY: CDS
: LOCATION: 252..1367
: US-07-813-584A-1

```

```

Alignment Scores:
Pred. No.: 0.268 Length: 1560
Score: 88.50 Matches: 37
Percent Similarity: 40.56% Conservative: 21
Best Local Similarity: 25.87% Mismatches: 46
Query Match: 11.39% Indels: 39
DB: 1 Gaps: 5

```

US-09-847-539A-6 (1-159) x US-07-813-584A-1 (1-1560)

```

QY 17 LeuThAsnLeuLeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArgAla 36
DB 984 GTTACATCTGAGTTAAACACAGCAAAAGCTCAACTCTCAGCGCTTGACAGAAAGTAGAAA 1043
QY 37 TleAspGluLeuLysGlnAlaIleGluAspLysGluAlaThrThrAlaIleGluAla 56
DB 1044 TTATCGAAAAAGAAAGAAAGCTGAGTTACAGCAAAA-----CTAGATGCA 1088
QY 57 AlaSerSerAspAlaLeuGlnAlaLeuAlaAspGlnThrAspAlaLeuGlnInserGlu 76
DB 1089 CAAGCAAAAGCCCTCAAGACACATTAGCAAAACAAACT-----GACAG 1133
QY 77 AlaAlaValValLysAlaAspAsnAlaAlaSerAspAlaLeuGlnAlaLeuAlaAspGln 96
DB 1134 CTTGCAAAAGCTTAAGACGCGGCGGCGGCTTCAAAAACACCTGCT----- 1181
QY 97 ThrAspAlaLeuGlnInserGluGlnAlaGluValValGlnSerAspAsnAlaAlaSerAsp 116
DB 1182 -----ACCAACACGCTAATTAAGAAAGCTCAGTCAAGCTCTCTCAAGCA 1226
QY 117 AlaTPGluLysAlaAlaThrProIleAlaLeuAspValLysThrLysAspThrLys 136
DB 1227 GCT-----ACNACA 1235
QY 137 ProValValLysGluGlnArgGlnAsnValAsnThrLeuProThrThrGluGlnGlu 156
DB 1236 CTTACCAAAATTAAGAAATGAGAG-----TCACATTTACCGTCAACAGCGCAAGCA 1286
QY 157 SerAsnPro 159
DB 1287 CCTAACCCA 1295

```

RESULT 39  
US-08-330-515-1  
; Sequence 1, Application US/08330515

```

: Patent No. 5556944
: GENERAL INFORMATION:
: APPLICANT: Fischetti, Vincent A.
: APPLICANT: Bessen, Debra E.
: TITLE OF INVENTION: No. 5556944e1 Immunoglobulin A Binding Protein
: NUMBER OF SEQUENCES: 3
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Saliwanchik & Saliwanchik
: STREET: 2421 N.W. 41st Street, Suite A-1
: CITY: Gainesville
: STATE: FL
: COUNTRY: USA
: ZIP: 32606
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/330,515
: FILING DATE:
: CLASSIFICATION: 530
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/813,584
: FILING DATE: 24-DEC-1991
: CLASSIFICATION: 530
: ATTORNEY/AGENT INFORMATION:
: NAME: Saliwanchik, David R.
: REGISTRATION NUMBER: RU-100.D1
: REFERENCE/DOCKET NUMBER: 31,794
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 904-372-5800
: TELEFAX: 904-375-8100
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1560 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: unknown
: MOLECULE TYPE: DNA (genomic)
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 2..40
: NAME/KEY: CDS
: LOCATION: 252..1367
: US-08-330-515-1

```

```

Alignment Scores:
Pred. No.: 0.268 Length: 1560
Score: 88.50 Matches: 37
Percent Similarity: 40.56% Conservative: 21
Best Local Similarity: 25.87% Mismatches: 46
Query Match: 11.39% Indels: 39
DB: 1 Gaps: 5

```

US-09-847-539A-6 (1-159) x US-08-330-515-1 (1-1560)

```

QY 17 LeuThAsnLeuLeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArgAla 36
DB 984 GTTACATCTGAGTTAAACACAGCAAAAGCTCAACTCTCAGCGCTTGACAGAAAGTAGAAA 1043
QY 37 TleAspGluLeuLysGlnAlaIleGluAspLysGluAlaThrThrAlaIleGluAla 56
DB 1044 TTATCGAAAAAGAAAGAAAGCTGAGTTACAGCAAAA-----CTAGATGCA 1088
QY 57 AlaSerSerAspAlaLeuGlnAlaLeuAlaAspGlnThrAspAlaLeuGlnInserGlu 76
DB 1089 CAAGCAAAAGCCCTCAAGACACATTAGCAAAACAAACT-----GACAG 1133
QY 77 AlaAlaValValLysAlaAspAsnAlaAlaSerAspAlaLeuGlnAlaLeuAlaAspGln 96
DB 1134 CTTGCAAAAGCTTAAGACGCGGCGGCGGCTTCAAAAACACCTGCT----- 1181

```

```

QY 97 ThrAspAlaLeuGlnSerGluGluValIaGlnSerAspAsnAlaLeuSerAsp 116
Db 1182 -----ACCAACACGCTAATAAGAAAGATCAGGTGCTCAACA 1226
QY 117 AlaTrpGluLysAlaAlaThrProIleAlaLeuAspValLysThrLysAspThrLys 136
Db 1227 GCT-----ACAGA 1235
QY 137 ProValValLysGluGluGlnAsnValAsnThrLeuProThrThrGluGlu 156
Db 1236 CCTAGCCAAATAAGAAATGAGA-----TCACAAATACGTCACAGCGCAAGA 1286
QY 157 SerAsnPro 159
Db 1287 GCTAACCA 1295

RESULT 40
US-08-533-669A-7
; Sequence 7, Application US/08533669A
; Patent No. 5834592
; GENERAL INFORMATION:
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE
; TITLE OF INVENTION: THERAPY AND DIAGNOSIS OF LEISHMANIASIS
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/533,669A
; FILING DATE: 22-SEP-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.420
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1771 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1698
; US-08-533-669A-7

Alignment Scores:
Pred. No.: 0.316 Length: 1771
Score: 88.50 Matches: 47
Percent Similarity: 45.34% Conservative: 26
Best Local Similarity: 29.19% Mismatches: 70
Query Match: 11.39% Indels: 18
DB: 2 Gaps: 6

US-09-847-539A-6 (1-159) x US-08-533-669A-7 (1-1771)
QY 8 ProArgIleIleProAsnGlyGlyThrLeuThrAsnLeuLeuGlyAsnAlaProGlyIu 27
Db 6 CCGGCTCCAGCGCCCTCGAGAGGAGCGGCTCTCCGCGGAGCTGGAGGCGGCGGAGGA 65

```

```

QY 27 sLeuAlaLeuArgAsnGluGluArgAlaIleAspGluLeuLysGluAlaIleIle 47
Db 66 GCGGCGCCGCTGGATGTCTATGCTAGCGCGCGAGGCCGCTGTCAGGCCCTCGAGA 125
QY 47 p-----LysGluAlaThrThrAlaIleGluAlaLeuSerSerAspAlaLe 62
Db 126 GGCAGCGCGCTCCGCGGAGCTGGAGAGGCCGAGAGAGGCCCGCTGATGTCTAT 185
QY 62 uGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaValLysAl 82
Db 186 GCATGCG---GCCGAGCAGGCCCGCTCAGGCCCTCGAGAGGAGCAGCGCTCCGCC 242
QY 82 aAsp-----AsnAlaAlaSerAspAla-----LeuGluAlaLeu-----AlaAspGln 97
Db 243 GAGCTGGAGGCTGCGGAGGAGGCGCGCTGAGGAGGAGGAGGAGGAGGAGGAGG 302
QY 97 rAspAlaLeuGlnSerGluGluValIaGlnSerAspAsnAlaLeuSerAspAl 117
Db 303 CCGCTCCAGCGCCCTCGAGAGGAGCGGCTCTCCGCGGAG----- 345
QY 117 aTrpGluLysAlaAlaThrProIleAlaLeuAspValLysThrLysAspThrLysPr 137
Db 346 -CTGAGAGAGAGCGAGAGGCGCGCCGCTGATGTCATGCGCGAGAGAGGCGCG 404
QY 137 oValValLysGluGluGlnAsnValAsnThrLeuProThrThrLysGluGlu 157
Db 405 CGTCCAGCGCCCTCGAGAGGAGCGGCTCTCCGCGGAGCTGGAGAGGCGGAGGAG 464
QY 157 r 157
Db 465 G 465

```

Search completed: October 13, 2002, 06:11:42  
 Job time : 67.016 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 13, 2002, 04:48:12 : Search time 224.471 Seconds

(without alignments)  
1216.148 Million cell updates/sec

Title: US-09-847-539A-6

Perfect score: 777

Sequence: 1 VDSPIEQPRIPNGTILTNL.....KKERQNVNLTPTTGESNP 159

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5

Ygapop 10.0 , Ygapext 0.5

Fgapop 6.0 , Fgapext 7.0

Delop 6.0 , Delext 7.0

Searched: 1736436 segs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL-frame+P2n.model -DEV-xlh

-O/cg2\_1/USPTO.spool/US09847539/runtat\_10102002\_093105\_4982/app-query.fasta.1.526

-DB-N.Geneseq\_032802 -QFMT-fastap -SUFFIX-stid.rng -MINMATCH=0.1 -LOOPCL=0

-LOOPEXT=0 -UNITS-bits -START=1 -END=-1 -MATRIX-blosum62 -TRANS-human40.ccd1

-LIST=45 -DOCALLIGN=200 -THR SCORE-pct -THR MAX=100 -THR MIN=0 -ALIGN=40

-MODE-LOCAL -OUTFMT=ptc -NORM-ext -HEAPSIZE=500 -MNLBN=0 -MAXLEN=2000000000

-USER-US09847539\_8CGN\_1.1.48-enumat\_10102002\_093105\_4982 -NCPU=6 -ICPU=3

-NO\_XLPY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV-TIMEOUT=120

-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FCGAPOP=6 -FCGAPEXT=7

-YGAPOP=10 -YGAPEXT=0.5 -DELPEXT=7

Database :

N\_Geneseq\_032802:\*

1: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1980.DAT:\*

2: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1981.DAT:\*

3: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1982.DAT:\*

4: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1983.DAT:\*

5: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1984.DAT:\*

6: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1985.DAT:\*

7: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1986.DAT:\*

8: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1987.DAT:\*

9: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1988.DAT:\*

10: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1989.DAT:\*

11: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1990.DAT:\*

12: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1991.DAT:\*

13: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1992.DAT:\*

14: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1993.DAT:\*

15: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1994.DAT:\*

16: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1995.DAT:\*

17: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1996.DAT:\*

18: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:\*

19: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1998.DAT:\*

20: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1999.DAT:\*

21: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:\*

22: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:\*

23: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:\*

24: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	777	100.0	654	21 AAD00560	Streptococcus pyog
2	777	100.0	764	21 AAD00559	Streptococcus pyog
3	760	97.8	504	21 AAD00564	Streptococcus pyog
4	750	96.5	469	21 AAD00562	Streptococcus pyog
5	694	89.3	853	21 AAD00563	Streptococcus pyog
6	692	89.1	777	21 AAD00561	Streptococcus pyog
7	188	24.2	1949	11 AAO06018	Sequence encoding
8	188	24.2	1950	10 AAO10093	Protein G gene. S
9	188	24.2	1950	10 AAN91094	Protein G gene. S
10	188	24.2	1950	11 AAO06017	Sequence encoding
11	188	24.2	1950	12 AAO10001	Sequence encoding
12	188	24.2	1950	11 AAO06464	Streptococcus Prot
13	187	24.1	2383	11 AAO06019	Sequence encoding
14	187	24.1	2383	12 AAO10002	Sequence encoding
15	187	24.1	2383	15 AAO75036	Streptococcus Prot
16	135	19.9	1576	21 AAZ47931	Streptococcus stra
17	139	17.9	798	11 AAO91099	Protein G gene var
18	139	17.9	798	11 AAO06009	Sequence encoding
19	139	17.9	798	12 AAO10007	Type 4 GX7809 prot
20	139	17.9	798	15 AAO64648	IgG-binding Strept
21	116	14.9	501	21 AAO45707	Arabidopsis thalia
22	114.5	14.7	1555	16 AAO68080	S. dysgalactiae ma
23	114.5	14.7	1555	16 AAO89197	S. dysgalactiae ma
24	112	14.4	1469	8 AAN70757	Sequence encoding
25	111	14.3	2121	23 ABL20337	Arabidopsis thalia
26	110	14.2	3212	20 AAX13365	Enterococcus faeca
27	109.5	14.1	812	21 AAC38790	Arabidopsis thalia
28	109	14.0	1128	21 AAC46811	Arabidopsis thalia
29	108	13.9	4086	23 ABL11003	Drosophila melanog
30	108	13.9	6782	20 AAX11002	Enterococcus faeca
31	107.5	13.8	486	20 AAX13883	Enterococcus faeca
32	105.5	13.6	989	18 AAT91293	Rabbit enteropatho
33	105	13.5	627	16 AAO84886	Synthetic protein
34	105	13.5	2526	16 AAO89196	S. dysgalactiae m1
35	104.5	13.4	575	21 AAC45194	Arabidopsis thalia
36	104	13.4	7434	23 AAS52179	Staphylococcus aur
37	104	13.4	7437	23 AAS52232	Staphylococcus aur
38	103.5	13.3	6228	23 AAS52178	Staphylococcus aur
39	103.5	13.3	6561	23 AAS55178	Staphylococcus aur
40	103.5	13.3	6744	13 AAO29471	Extracellular fact
41	103	13.3	7104	23 AAS51998	Staphylococcus aur
42	103	13.2	7107	23 AAS54654	Staphylococcus aur
43	102.5	13.2	1005	23 AAS92880	DNA encoding novel
44	102	13.1	603	16 AAO84887	Chimeric syntheti
45	102	13.1	8155	18 AAV74374	Staphylococcus aur

## ALIGNMENTS

RESULT 1  
AAD00560  
ID AAD00560 standard; DNA; 654 BP.

XX AAD00560;

XX 29-AUG-2000 (first entry)

XX Streptococcus pyogenes strain SF370 GRAB protein coding region.

XX GRAB protein; protein G related alpha2m binding protein; vaccine;

XX alpha2-macroglobulin; group A Streptococcus; GAS; antiheliotic;

XX Immune response; Streptococcus pyogenes infection; ds.

XX Streptococcus pyogenes.

OS Key Location/Qualifiers  
FH CDS 1..654

```
FT      /*tag- a  
FT      /product= "GRAB protein"  
FT      1...99  
FT      sig-peptide  
FT      /*tag- b  
FT      mat-peptide  
FT      100..651  
FT      /*tag- c  
FT      /product= "Mature GRAB protein"  
PN      WO200026240-A2.  
XX      11-MAY-2000.  
XX      02-NOV-1999: 99WO-GB03631.  
XX      02-NOV-1998: 98GB-0023975.  
XX      (ACTI-) ACTINOVA LTD.  
XX      Bjorck LH, Rasmussen M;  
XX      WPI: 2000-365572/31.  
XX      P-PSDB: AAY71042.  
XX      New alpha2m binding protein for generating a protective immune response  
XX      to group A streptococcus and purifying the binding protein  
XX      Claim 13; Page 63; 67pp; English.  
XX      The patent discloses a new family of proteins termed GRAB (protein G  
XX      related alpha2m binding protein) from Streptococcus pyogenes which have  
XX      the ability to bind alpha2-macroglobulin (alpha2m) and show homology to  
XX      protein G of group G Streptococcus. GRAB protein and peptides derived  
XX      from it are used in vaccine compositions for generating a protective  
XX      immune response against group A Streptococcus. Antibodies against GRAB  
XX      are useful for treating Streptococcus pyogenes infections. The protein  
XX      is also useful for purifying Streptococcus pyogenes strain SF370.  
XX      is a DNA encoding full-length GRAB protein from S. pyogenes strain SF370.  
XX      Sequence 654 BP; 234 A; 118 C; 143 G; 159 T; 0 other;  
XX      Alignment Scores:  
XX      Pred. No.: 1 09e-74 Length: 654  
XX      Score: 777.00 Matches: 159  
XX      Percent Similarity: 100.00% Conservative: 0  
XX      Best Local Similarity: 100.00% Mismatches: 0  
XX      Query Match: 100.00% Indels: 0  
XX      DB: 21 Gaps: 0  
US-09-847-539A-6 (1-159) x AAD00560 (1-654)  
QY      1 ValAspSerProIleGluInProArgIleIleProAsnGlyThrLeuThrAsnLeu 20  
DB      100 GTTGACCTCACTCGACAGCGCTCGAATTATTCCTCAATGGCGAACCTTAACCTAATCTT 159  
QY      21 LeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluLysArgAlaLeu 40  
DB      160 CTGGCAATCTCTCAGAAAACCTGCATTACATGAAGAAACAGCCATTGATGATTA 219  
QY      41 LysLysGluAlaIleGluAspLysGluAlaIleThrAlaIleGluAlaIleAspGluLeu 60  
DB      220 AAAAACAAGCTATGAGGATTAAGAAGCTACAGCATATTAACACAGCAAGCTTCAGAT 279  
QY      61 AlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaIleVal 80  
DB      280 GCCTTAGAAGCATTTACCGGATCAACAGACGCTTTACATCAAGAAAGCTGCGTTGTT 339  
QY      81 LysAlaAspAsnAlaIleAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeu 100  
DB      340 AAAGCGGATAACGCTCTAGTACGCGCTTAGAAGCATTTGCGGATCAACAGACGCTTTA 399  
QY      101 GlnSerGluGluAlaGluAlaValGlnSerAspAsnAlaIleAspAlaIleProGluLys 120  
DB      400 CAATCAGAAAGAGCTAAGTAGTCAATCAGATTAACGCTGCTAGTGACGCTGGAAAAA 459
```

```
QY      121 AlaIleThrProIleAlaLeuAspValLysLysThrLysAspThrLysProValIleVal 140  
DB      460 GCACCAACTCCAAATCGCTTTAGATGTTAAGAAAACCTAAAGTACAAAACCTGATGTTAAA 519  
QY      141 LysGluGluArgGlnAsnValAsnThrLeuProThrThrGluGluSerAsnPro 159  
DB      520 AAAGAGAAAGACAAAACGTTAATACCTTCCTACAACTGGTGAAGAGTCTAACCCA 576  
RESULT 2  
AAD00559  
ID      AAD00559 standard; DNA: 764 BP.  
AC      AAD00559;  
XX      29-AUG-2000 (first entry)  
XX      Streptococcus pyogenes strain SF370 GRAB protein encoding DNA.  
XX      GRAB protein: protein G related alpha2m binding protein; vaccine;  
XX      alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;  
XX      immune response; Streptococcus pyogenes infection; ds.  
XX      Streptococcus pyogenes.  
XX      Key Location/Qualifiers  
XX      CDS 60..713  
XX      /*tag- a  
XX      /product= "GRAB protein"  
XX      60..158  
XX      /*tag- b  
XX      159..710  
XX      /*tag- c  
XX      /product= "Mature GRAB protein"  
XX      complement (101..124)  
XX      /tag- d  
XX      /bound_moiety= "Primer 1 represented in AAD00565"  
XX      complement (101..127)  
XX      /*tag- e  
XX      /bound_moiety= "Primer 2 represented in AAD00566"  
XX      complement (160..184)  
XX      /*tag- f  
XX      /bound_moiety= "Primer 3 represented in AAD00567"  
XX      complement (563..594)  
XX      /*tag- g  
XX      /bound_moiety= "Primer 4 represented in AAD00568"  
XX      complement (605..626)  
XX      /*tag- h  
XX      /bound_moiety= "Primer 5 represented in AAD00569"  
XX      WO200026240-A2.  
XX      11-MAY-2000.  
XX      02-NOV-1999: 99WO-GB03631.  
XX      02-NOV-1998: 98GB-0023975.  
XX      (ACTI-) ACTINOVA LTD.  
XX      Bjorck LH, Rasmussen M;  
XX      WPI: 2000-365572/31.  
XX      P-PSDB: AAY71042.  
XX      New alpha2m binding protein for generating a protective immune response  
XX      to group A streptococcus and purifying the binding protein  
XX      Example 1; Fig 2B; 67pp; English.  
XX      The patent discloses a new family of proteins termed GRAB (protein G  
XX      related alpha2m binding protein) from Streptococcus pyogenes which have  
XX      the ability to bind alpha2-macroglobulin (alpha2m) and show homology to
```

CC protein G of group G Streptococcus. GRAB protein and peptides derived  
CC from it are used in vaccine compositions for generating a protective  
CC immune response against group A Streptococcus. Antibodies against GRAB  
CC are useful for treating Streptococcus pyogenes infections. The protein  
CC is also useful for purifying alpha2m from a sample. The present sequence  
CC is a DNA encoding full-length GRAB protein from S. pyogenes strain SF370.  
XX

Sequence 764 BP; 279 A; 131 C; 159 G; 195 T; 0 other;

## Alignment Scores:

Pred. No.:	1-31e-74	Length:	764
Score:	777.00	Matches:	159
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	21	Gaps:	0

US-09-847-539a-6 (1-159) x AAD00559 (1-764)

QY 1 ValAspSerProIleGluInProArgIleIleProAsnGlyGlyThrLeuThraSleu 20  
DB 159 GTTGACTCCACCTATCGAAGCCCTCGAATTAATTCCTCAATGCGGAACTTAATCTT 218  
QY 21 LeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaIleAspGlu 40  
DB 219 CTGGCAATGCTCCAGAAAACCTGCAATTCATGATGAGAAAGAGCCATTGATGATTA 278  
QY 41 LysLysGlnAlaIleGluAspLysGluAlaThrThraIleGluAlaIleAspSer 60  
DB 279 AAAAAACAACCTATTGAGTAAGAAAGCTAACGACCTATAGAAAGCAAGTTCCAGAT 338  
QY 61 AlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaIleVal 80  
DB 339 GCCCTTAAGACATTAGCGGATCAACAGACCTTACATCAGAAAGCTGCGGTGTT 398  
QY 81 LysAlaAspAsnAlaIleAspAlaLeuGlnAlaLeuAlaAspGlnThrAspAlaLeu 100  
DB 399 AAGCGGATTAACGCTGTAGAGACGCTTGAACATTCGGCGGATCAACAGAGCTTTA 458  
QY 101 GlnSerGluGluAlaGluValaGlnSerAspAsnAlaIleAspAlaIleAspGlu 120  
DB 459 CAATCAGAAAGAAAGCTGAAGTTCATTCAGATTAACCTGCTAGTGAAGCGGAAAAA 518  
QY 121 AlaIleThrProIleAlaLeuAspValLysThrLysAspThrLysProValValLys 140  
DB 519 GCACCAACTCCAACTGCTTAGATGTTAAGAAAGCTAAAGATACAAAACCTGTAGTTAA 578  
QY 141 LysGluGluAlaArgGlnAsnValaThrLeuProThrThrGlyGluGluSerAsp 159  
DB 579 AAAGCAAGAAAGAAAGCTTAATACCTTCCTCAAACTGCTGAAGTCTAACCA 635

## RESULT 3

AAD00564 standard; DNA; 504 BP.

AAD00564:

29-AUG-2000 (first entry)

Streptococcus pyogenes strain KTL3 partial GRAB protein encoding DNA.

GRAB protein; protein G related alpha2m binding protein; vaccine;

alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;

immune response; Streptococcus pyogenes infection; ds.

Streptococcus pyogenes.

Location/Qualifiers

Key 1..504  
FT /tag- a  
FT /product- "GRAB protein"  
FT /partial  
XX

PN WO200026240-A2.

XX 11-MAY-2000.

XX 02-NOV-1999; 99WO-GB03631.

XX 02-NOV-1998; 98GB-0023975.

XX (ACTI-) ACTINOVA LTD.

XX Bjorck LH, Rasmussen M;

XX WPI: 2000-365572/31.

XX P-PSDB; AAY71046.

XX New alpha2m binding protein for generating a protective immune response

XX to group A streptococcus and purifying the binding protein

XX Claim 13; Page 65; 67pp; English.

CC The patent discloses a new family of proteins termed GRAB (protein G  
CC related alpha2m binding protein) from Streptococcus pyogenes which have  
CC the ability to bind alpha2-macroglobulin (alpha2m) and show homology to  
CC protein G of group G Streptococcus. GRAB protein and peptides derived  
CC from it are used in vaccine compositions for generating a protective  
CC immune response against group A Streptococcus. Antibodies against GRAB  
CC are useful for treating Streptococcus pyogenes infections. The protein  
CC is also useful for purifying alpha2m from a sample. The present sequence  
CC is a DNA encoding partial GRAB protein from S. pyogenes strain KTL3.  
CC The protein has alpha2m binding region and is useful in vaccine  
CC composition.

XX Sequence 504 BP; 188 A; 97 C; 108 G; 111 T; 0 other;

## Alignment Scores:

Pred. No.:	5.32e-73	Length:	504
Score:	760.00 <td>Matches:</td> <td>156</td>	Matches:	156
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	97.81%	Indels:	0
DB:	21	Gaps:	0

US-09-847-539a-6 (1-159) x AAD00564 (1-504)

QY 1 ValAspSerProIleGluInProArgIleIleProAsnGlyGlyThrLeuThraSleu 20  
DB 34 GTTGACTCCACCTATCGAAGCCCTCGAATTAATTCCTCAATGCGGAACTTAATCTT 93  
QY 21 LeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaIleAspGlu 40  
DB 94 CTGGCAATGCTCCAGAAAACCTGCAATTCATGATGAGAAAGAGCCATTGATGATTA 153  
QY 41 LysLysGlnAlaIleGluAspLysGluAlaThrThraIleGluAlaIleAspSer 60  
DB 154 AAAAAACAACCTATTGAGTAAGAAAGCTGAAGCTATAGAAAGCAAGTTCCAGAT 213  
QY 61 AlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaIleVal 80  
DB 214 GCCTTAGAACCTATAGCGGATCAACAGACCTTACATCAGAAAGCTGCGGTGTT 273  
QY 81 LysAlaAspAsnAlaIleAspAlaLeuGlnAlaLeuAlaAspGlnThrAspAlaLeu 100  
DB 274 AAAGCGGATTAACGCTGTAGAGACGCTTGAACGATTCGGGATCAACAGAGCTTTA 333  
QY 101 GlnSerGluGluAlaGluValaGlnSerAspAsnAlaIleAspAlaIleAspGlu 120  
DB 334 CAATCAGAAAGAAAGCTGAAGTTCATCAGATTAACGCTGTAGTGAAGCGGAAAAA 393  
QY 121 AlaIleThrProIleAlaLeuAspValLysThrLysAspThrLysProValValLys 140  
DB 394 GCACCAACTCCAACTGCTTAGATGTTAAGAAAGCTAAAGATACAAAACCTGTAGTTAA 453  
QY 141 LysGluGluAlaArgGlnAsnValaThrLeuProThrThrGlyGluGlu 156

|||||  
DB 454 AAGAGAAAGACAAACGTTATACCTTCCTCAACAAGTGGAAG 501  
RESULT 4  
AAD00562  
ID AAD00562 standard; DNA: 469 BP.  
XX  
AC AAD00562;  
XX  
DT 29-AUG-2000 (first entry)  
XX  
XX Streptococcus pyogenes strain AP1 partial GRAB protein encoding DNA.  
DE  
XX GRAB protein; protein G related alpha2M binding protein; vaccine;  
KW alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;  
KM Immune response; Streptococcus pyogenes infection; ds.  
XX  
OS Streptococcus pyogenes.  
XX  
FH Key Location/Qualifiers  
FT CDS 1..468  
FT /tag= a  
FT /product= "GRAB protein"  
FT /partial  
XX  
XX WO200026240-A2.  
XX  
XX 11-MAY-2000.  
XX  
XX 02-NOV-1999; 99WO-GB03631.  
XX  
XX 02-NOV-1998; 98GB-0023975.  
XX  
XX (ACTI-) ACTINOVA LTD.  
XX  
XX Bjorck LH, Rasmussen M:  
XX  
XX WPI: 2000-365572/31.  
XX  
XX P-PSDB: AAY71044.  
XX  
XX New alpha2M binding protein for generating a protective immune response  
XX to group A streptococcus and purifying the binding protein -  
XX  
XX  
XX Claim 13; Page 64; 67pp; English.  
XX  
XX The patent discloses a new family of proteins termed GRAB (protein G  
XX related alpha2M binding protein) from Streptococcus pyogenes which have  
XX the ability to bind alpha2-macroglobulin (alpha2M) and show homology to  
XX protein G of group G Streptococcus. GRAB protein and peptides derived  
XX from it are used in vaccine compositions for generating a protective  
XX immune response against group A Streptococcus. Antibodies against GRAB  
XX are useful for treating Streptococcus pyogenes infections. The protein  
XX is also useful for purifying alpha2M from a sample. The present sequence  
XX is a DNA encoding partial GRAB protein from S. pyogenes strain AP1.  
XX  
XX The protein has alpha2M binding region and is useful in vaccine  
XX composition.  
XX  
XX Sequence 469 BP; 180 A; 90 C; 99 G; 100 T; 0 other;  
XX  
XX  
XX Alignment Scores:  
XX Pred. No.: 5.79e-72 Length: 469  
XX Score: 750.00 Matches: 154  
XX Percent Similarity: 99.35% Conservative: 0  
XX Best Local Similarity: 99.35% Mismatches: 1  
XX Query Match: 96.53% Indels: 0  
XX DB: 21 Gaps: 0  
XX  
XX US-09-847-539a-6 (1-159) x AAD00562 (1-469)  
XX  
XX 2 AppserProileglnuInproArglelleleproAsnglGlyThreuleuThrasnleu 21  
XX |  
DB 1 GACTCACTATGAGACAGCGCTAGAAATTATTCACAAATGGCGAAGCTTAATTATCTTCTT 60

QY 22 GlyAsnAlaProGluLysIleuAlaLeuArgAsnGluGluArgAlaIleAspGluLeuLys 41  
DB 61 GGCAATGCTCCACAGAAAACATGGCATTACGTAATGAAGAAAAGACCATTGATGAATTTAAAA 120  
QY 42 LysGlnAlaIleGluAspLysGluAlaThrThrAlaIleGluAlaIleSerAspAla 61  
DB 121 AAACAAGCTATTGAGATTAAGAAAGCTACACAGCTATAGAACACAGCAAGTTGATGATGCC 180  
QY 62 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaIleValLys 81  
DB 181 TTAGAAAGCATTTAGCGGATCAACAGACGCTTTCAATTCAGAAACAGCTGGTGTAAAA 240  
QY 82 AlaAspAsnAlaIleAserAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGln 101  
DB 241 GCGGATTAACGCTGCTAGTGACGCCCTTAGAAGCATTTGGCGGATCAACAGACGCTTTACAA 300  
QY 102 SerGluGluAlaGluValValGlnSerAspAsnAlaIleSerAspAlaIleTrpGluLysAla 121  
DB 301 TCAGAAAGAGCTGAGTAGTTCATTCAGATTAACGCTAGTACGCCCTGGGAAAAAGCA 360  
QY 122 AlaThrProIleAlaLeuAspValLysLysThrLysAspThrLysProValValLysLys 141  
DB 361 GCACCTCCAACTCCCTTTAGATGTTAAGAAACTTAAGATTAACAACCTGTAGTTAAAAA 420  
QY 142 GluGluArgGlnAsnValAsnThrLeuProThrThrGlyGluGlu 156  
DB 421 GAAGAAAGACAAACGTTAATACCTTCCTACAACTGGTGAAGAG 465  
RESULT 5  
AAD00563  
ID AAD00563 standard; DNA: 853 BP.  
XX  
XX AAD00563;  
XX  
XX 29-AUG-2000 (first entry)  
XX  
XX Streptococcus pyogenes strain AP49 partial GRAB protein encoding DNA.  
XX  
XX  
XX Streptococcus pyogenes strain AP49 partial GRAB protein encoding DNA.  
XX  
XX GRAB protein; protein G related alpha2M binding protein; vaccine;  
XX alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;  
XX Immune response; Streptococcus pyogenes infection; ds.  
XX  
XX Streptococcus pyogenes.  
XX  
XX  
XX Key Location/Qualifiers  
XX FT CDS 37..852  
XX /tag= a  
XX /product= "GRAB protein"  
XX /partial  
XX  
XX WO200026240-A2.  
XX  
XX 11-MAY-2000.  
XX  
XX 02-NOV-1999; 99WO-GB03631.  
XX  
XX 02-NOV-1998; 98GB-0023975.  
XX  
XX (ACTI-) ACTINOVA LTD.  
XX  
XX Bjorck LH, Rasmussen M:  
XX  
XX WPI: 2000-365572/31.  
XX  
XX P-PSDB: AAY71045.  
XX  
XX New alpha2M binding protein for generating a protective immune response  
XX to group A streptococcus and purifying the binding protein -  
XX  
XX  
XX Claim 13; Page 64; 67pp; English.  
XX  
XX The patent discloses a new family of proteins termed GRAB (protein G  
XX related alpha2M binding protein) from Streptococcus pyogenes which have  
XX the ability to bind alpha2-macroglobulin (alpha2M) and show homology to



CC protein G of group G Streptococcus. GRAB protein and peptides derived  
 CC from it are used in vaccine compositions for generating a protective  
 CC immune response against group A Streptococcus. Antibodies against GRAB  
 CC are useful for treating Streptococcus pyogenes infections. The protein  
 CC is also useful for purifying alpha2M from a sample. The present sequence  
 CC is a DNA encoding partial GRAB protein from S. pyogenes strain AP49.  
 CC The protein has alpha2M binding region and is useful in vaccine  
 CC composition.

XX Sequence 853 BP; 295 A; 171 C; 197 G; 190 T; 0 other;

#### Alignment Scores:

Pred. No.:	1,25e-65	Length:	853
Score:	694.00	Matches:	156
Percent Similarity:	58.21%	Conservative:	0
Best Local Similarity:	58.21%	Mismatches:	0
Query Match:	89.32%	Indels:	112
DB:	21	Gaps:	1

US-09-847-539a-6 (1-159) x AAD00563 (1-853)

```

QY 1 ValAspSerProIleGluInProArgIleIleProAsnGlyGlyThrLeuThrAsnLeu 20
DB 46 GTTGACTCACCCTATCGAAGACGCTCGAATATTCGAATGCGGACCTTAACCTAATCTT 105
QY 21 LeuGlyAsnAlaProGlyLysLeuAlaLeuArgAsnGluLysArgAlaIleAspGluLeu 40
DB 106 CTTGGCAGATCTCCAGAAAAACGTGCAATGATGAAGAAAGAGGCAATGATGATTA 165
QY 41 LysLysGlnAlaIleGluAspLysGluAlaThrAlaIleGluAlaIleAsp----- 58
DB 166 AAAAAACAACCTATTGAGGATTAAGAAAGCTACACACCTATAGACAGCAAGCTTCAGAT 225
QY 58 ----- 58
DB 226 GCCTTAGAAGCATTAGCGGATCAAGACGCTTACAAATCAGAAAGAGCTGCAGTAGTT 285
QY 58 ----- 58
DB 286 CAATCAGATTAACGCTGCTAGTACGCCCTTAGAAGCATTAGCGGATCAAGACGCTTTA 345
QY 58 ----- 58
DB 346 CAATCAGAAGAGAGCTGAGTAGTTCATCAGATTAAGCCTGCTGCTGACGCCCTTAGAAGCA 405
QY 58 ----- 58
DB 406 TTGGCGGATCAAAACAGACGCTTACAAATCAGAAAGAGCCTTCGTTTAAAGCGATTAAC 465
QY 58 ----- 58
DB 466 GCTGCTAGTACAGCCTTAGAAGCATTGGCGGATCAAAACAGACGCTTTCACATCAAGAA 525
QY 59 -----SerAspAlaLeuGluAlaIleAlaLeuAlaIleAspGln 68
DB 526 GCTTCGGTGTGTAAGGAGTAACGCTGCTAGTACGCCCTTAGAAGCATTGGCGGATCA 585
QY 69 ThrAspAlaLeuGlnSerGluGluAlaIleAlaValAlaLysAlaAspAlaIleSerAsp 88
DB 586 ACAGAGCCTTACAAATCAGAAAGAGCTGCTGTTAAAGCGATTAACGCTGCTAGTAGAC 645
QY 89 AlaLeuGluAlaLeuAlaIleAspGlnThrAspAlaLeuGlnSerGluGluAlaIleValAla 108
DB 646 GCCCTTAAGACATTGGCGGATCAAAACAGACGCTTACAAATCAGAAAGACCTGAAGTAGTT 705
QY 109 GluSerAspAlaAlaIleSerAspAlaIleArgLysAlaIleAlaIleThrProIleAlaLeuAsp 128
DB 706 CAATCAATATAACGCTGCTAGTACGCCCTGGGAAAAAGCAGCAACTCCAATGCTTTAGAT 765
QY 129 ValLysLysThrLysAspThrLysProValValLysLysGluLysGluArgGlnAsnValAsn 148
DB 766 GTTAAAGAAACCTAAAGATACAAAACCTGTAGTTAAAAAGAAAGAAACAAACGTTAAT 825
  
```

QY 149 ThrLeuProThrArgLysGluIn 156  
 DB 826 ACCCTTCTACAACTGCTGAAGAG 849

#### RESULT 6

AAD00561  
 ID AAD00561 standard; DNA: 777 BP.

AC AAD00561;

DT 29-AUG-2000 (first entry)

DE Streptococcus pyogenes strain KTL9 partial GRAB protein encoding DNA.

KW GRAB protein; protein G related alpha2M binding protein; vaccine;

KW alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;

KW Immune response; Streptococcus pyogenes infection; ds.

OS Streptococcus pyogenes.

PH Key Location/Qualifiers

FT CDS 1..777

FT /lag- a

FT /product- "GRAB protein"

FT /note- "Does not include stop codon"

PN WO200026240-A2.

PD 11-MAY-2000.

PF 02-NOV-1999; 99WO-GB03631.

PR 02-NOV-1998; 98GB-0023975.

PA (ACT1-) ACTINOVA LTD.

PI Bjorck LH, Rasmussen M;

DR WPI; 2000-365572/31.

PS P-PSDB; AAY71043.

PT New alpha2M binding protein for generating a protective immune response

PT to group A streptococcus and purifying the binding protein

PS Claim 13; Page 63-64; 67pp; English.

CC The patent discloses a new family of proteins termed GRAB (protein G

CC related alpha2M binding protein) from Streptococcus pyogenes which have

CC the ability to bind alpha2-macroglobulin (alpha2M) and show homology to

CC protein G of group G Streptococcus. GRAB protein and peptides derived

CC from it are used in vaccine compositions for generating a protective

CC immune response against group A Streptococcus. Antibodies against GRAB

CC are useful for treating Streptococcus pyogenes infections. The protein

CC is a DNA encoding partial GRAB protein from S. pyogenes strain KTL9.

CC The protein has alpha2M binding region and is useful in vaccine

CC composition.

SO Sequence 777 BP; 269 A; 151 C; 178 G; 179 T; 0 other;

#### Alignment Scores:

Pred. No.:	1,83e-65	Length:	777
Score:	692.00	Matches:	153
Percent Similarity:	64.02%	Conservative:	0
Best Local Similarity:	64.02%	Mismatches:	2
Query Match:	89.06%	Indels:	84
DB:	21	Gaps:	1

US-09-847-539a-6 (1-159) x AAD00561 (1-777)

```

QY 1 ValAspSerProIleGluInProArgIleIleProAsnGlyGlyThrLeuThrAsnLeu 20
  
```

```

Db 61 GTTGAACCTACCTATCGACAGCCCTCGAATTATTCCAAATGCCGCAACCTTAATCTT 120
Qy 21 LeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaIleAspGluLeu 40
Db 121 CTTGGCAATGCTCCAGAAAACTGGCATTTACGTAATGAAGAAAGGCCCTTGTGTAATTA 180
Qy 41 LysLysGlnAlaIleGluAspLysGluAlaIleThrAlaIleGluAlaIleAspSerAsp 60
Db 181 AAAAAACAAGCTATTGAGGATTAAGAGAGCTACGACAGCTATATGAGACAGCAAGTTCCAGAT 240
Qy 61 Ala----- 61
Db 241 GCCTTAGAAGATTAGCGGATCAAGCAGACGCTTTACAATCAGAGACAGCTGCATGATT 300
Qy 61 ----- 61
Db 301 CAATCAGATTAAGCTGCTAGTAGAGCGCTTAGAAGCATTTGGCGGATCAACAGACGCTTAA 360
Qy 61 ----- 61
Db 361 CAATCAGAAAGCTGCGTGTGTAAAGCGGATTAACGCTGTAGTGACACTTTAGAAAGCA 420
Qy 61 ----- 61
Db 421 TTGGCGGATCAAAACAGACCGCTTTACATCAGAAAGAGCTGCGTGTAAAGCGGATTAAC 480
Qy 62 -----LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGlu 76
Db 481 GCTGCTAGTAGACCTTTAAAGCATTTGGCGGATCAACAGACGCTTTACATCAGAAAGAA 540
Qy 77 AlaAlaValAlaIleAlaAspAsnAlaIleSerAspAlaLeuGluAlaLeuAlaAspGln 96
Db 541 GCTGCGGTTGTTAAAGCGGATTAACGCTGTAGTGACACTTTAGAAAGCATTTGGCGGATCAA 600
Qy 97 ThrAspAlaLeuGlnSerGluGluAlaIleValGlnSerAspAsnAlaIleSerAsp 116
Db 601 ACAGACGCTTTACATCAGAAAGAGCTGAAGTACTCAATCAGATTAACCTGCTAGTGAC 660
Qy 117 AlaThrGluLysAlaIleAlaThrProIleAlaLeuAspValLysLysThrLysAspThrLys 136
Db 661 GCCTGGGAGAAAGACAGCACTCCATCGCTTAGATGTTAAGAAAACTAAAGATACAAAA 720
Qy 137 ProValValLysLysGluGluArgGlnAsnValAsnThrLeuProThrThrGluGlu 155
Db 721 CCTGTAGTTAAAAAGAGAAAGACAAAAAGCTTAATACCTCTCTCAACAGTGTGAA 777
RESULT 7
AA006018
ID AA006018 standard: DNA; 1949 BP.
XX
AC AA006018;
XX
DT 17-JUN-1991 (first entry)
XX
DE Sequence encoding Protein G with two active sites.
XX
KW Immunoglobulin: ds.
XX
OS Streptococcus sp. Lancefield Group G strain.
XX
FH
FT Key Location/Qualifiers
FT CDS 577..1920
FT FT /*tag= a
FT FT -35_signal 465..470
FT FT /*tag= b
FT FT -10_signal 487..492
FT FT /*tag= c
FT FT RBS 565..571
FT FT /*tag= d
XX
XX USA956296-A.
XX
XX 11-SEP-1990.
PD

```

```

XX XX
PF 20-JUN-1988; 88US-0209236.
XX XX
PR 20-JUN-1988; 88US-0209236.
PR 14-FEB-1986; 86US-0829354.
PR 23-APR-1986; 86US-0854887.
PR 17-FEB-1987; 87MO-US00329.
PR 19-JUN-1987; 87US-0063959.
XX XX
PA (GENE-) GENEX CORP.
XX
PI Fahnstock SR;
XX
XX WPI: 1990-297491/39.
DR P-PSDB; AAR07013.
XX
XX PT Recombinant Protein G variants - obt'd. using a cloned gene
PT encoding Protein G from Streptococcus sp., used for binding
PT Immunoglobulin.
XX
XX PS Disclosure: Fig 8a-c; 48pp; English.
XX
XX CC Fragments and variants of the sequence are claimed esp. where
XX CC incorporated into a non-pathogenic host eg. E.coli, and expressed
XX CC at high levels.
XX CC The variants have a higher binding efficiency and capacity for
XX CC Immunoglobulin, and may be used for purifying, detecting and
XX CC isolating antibodies.
XX
SQ Sequence 1949 BP; 705 A; 323 C; 398 G; 523 T; 0 other;

Alignment Scores:
Pred. No.: 8,12e-11 Length: 1949
Score: 188.00 Matches: 61
Percent Similarity: 51.59% Conservative: 20
Best Local Similarity: 38.85% Mismatches: 60
Query Match: 24.20% Indels: 16
DB: Gaps: 6

US-09-847-539A-6 (1-159) x AA006018 (1-1949)
Qy 1 ValAspSerProIleGluGlnProArgIleIleProAsnGlyLysThrLeuThrAsnLeu 20
Db 676 GTTGATTCACCAATCGAATACCCCAATTAATCGTAAGAGAGTGTGTAATTAATCTT 735
Qy 21 LeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaIleAspGluLeu 40
Db 736 CTGGGAAATTCAGACAGCAACTGCGCTTGGCTTAATGAAGAGAGTGTCTACAGCTGATTG 795
Qy 41 LysLysGlnAlaIleGluAspLysGluAlaIleThrAlaIleGluAlaIleAspSerAsp 60
Db 796 ACAGCAGCAGCGGTAGCCGATCTGTCGACAGCGCAGCGGAGAAATGCTGGGCGAGCA 855
Qy 61 AlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaIleVal 80
Db 856 GCTTGGGAGAGCGGCA--CCAGCAGATGCTCTTAGCAAAAGCCAAACAGATGCCCTT 912
Qy 81 LysAlaAspAsn-----AlaAlaSerAspAlaLeuGluAlaLeu-----AlaAsp 95
Db 913 AAAAGATTCAACAAATATGAGTAGTAAGTACTATTACAAATCTAATCAACATGCCAA 972
Qy 96 GlnThrAspAlaLeuGlnSerGluGluAlaIleValGlnSerAspAsnAlaIle--- 114
Db 973 ACTGTGAAGCATTAAGAACCTTCAAGCACAAGTGTGTAATCAGCAGAAAGCGCGT 1032
Qy 115 ---SerAspAlaIleGlu-----LysAlaAlaIleThrProIleAlaLeu 127
Db 1033 ATTTCAGAAAGCAACAGATGCTTATCTGATTCTTCTGAAATCCCAACACCTGCTGAAGAT 1092
Qy 128 AspValLysLysThrLys-----AspThrLysProValValLysGlu 142
Db 1093 ACTGTTAAATCAATTAATTAAGTACGTAAGCTTAAGCTTACGTAAGAGAA 1143

```

```

RESULT 8
AAN91093
ID AAN91093 standard; DNA: 1950 BP.
XX
AC AAN91093;
XX
DT 04-JUL-1990 (first entry)
XX
DE Protein G gene.
XX
KW Protein G; immunoglobulin; Fc receptor; ds.
XX
OS Streptococcus sp.
XX
FH Key Location/Qualifiers
FT -35_signal /+tag- c 465..470
FT -10_signal 487..492 /+tag- b
FT RBS 565..571 /+tag- d
FT CDS 578..1918 /*tag- a
FT
FT
XX
PN MO8810306-A.
XX
PD 29-DEC-1988.
XX
PF 20-JUN-1988; 88WO-US02084.
XX
PR 19-JUN-1987; 87US-0063959.
XX
PA (GENE-) GENEX CORP.
XX
PI Fahnestock SR;
XX
DR WPI: 1989-023848/03.
DR P-PSDB; AAP95030.
XX
PT Cloned protein G variant genes -
PT expressing proteins having immunoglobulin-binding properties of
PT protein G and derived from Streptococcus sp.
XX
PS Disclosure; ; 11pp; English.
XX
CC Gene encodes protein G of non-pathogenic streptococcus sp. allowing
CC isolation of the protein and variants, useful as bacterial Fc receptors
CC eg in purification and detection of Abs., screening of hybridoma clones
CC and treatment of disease.
XX
SQ Sequence 1950 BP; 705 A; 323 C; 398 G; 524 T; 0 other;

Alignment Scores:
Pred. No.: 8,12e-11 Length: 1950
Score: 188.00 Matches: 61
Percent Similarity: 51.59% Conservative: 20
Best Local Similarity: 38.85% Mismatches: 60
Query Match: 24.20% Indels: 16
DB: Gaps: 6

US-09-847-539a-6 (1-159) x AAN91093 (1-1950)
OY 1 ValasparProIleGluInProArgIleIleProAsnGlyIleThrLeuThrAsnLeu 20
DB 677 GTTGATTCACCAATCGAAGATACCCCAATATTCGTAATGCTGCGAATTAACCTAATCTT 736
OY 21 LeuLysAlaIleProGluLysLeuAlaLeuArgAsnGluGluAlaIleAspGluLeu 40
DB 737 CTGGGGAATTCAGACAAACACTGCTTGGTATGAGAGAGCTGACAGCTGATTG 796
OY 41 LysLysGlnAlaIleGluAspLysGluAlaThrThrAlaIleGluAlaIleSerSerasp 60
DB 797 ACACAGCAGCGGCTAGCCGATACCTGCTGACAGCAGCGGCTGAAATGCTGGGCGACAGA 856

```

```

OY 61 AlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuInserGluGluAlaAlaVal 80
DB 857 GCTTGGAGACGCGCA--GCAGCAGATGCTTACGCAAAAGCCAAAGATGCCCTT 913
OY 81 LysAlaAspAsn-----AlaAlaSerAspAlaLeuGluAlaLeu-----Alaasp 95
DB 914 AAGCAATTCACCAATATGAGTGAAGTACTATTACAGAACTTAATCAACATGCCAA 973
OY 96 GlnThrAspAlaLeuInserGluGluAlaGluValAlaInserAspAsnAlaAla 114
DB 974 ACTGTTGACGCATAAAGACCTTCAGACACAGTGTGATCAGACGAGAAAGCCGT 1033
OY 115 ---SerAspAlaTrpGlu-----LysAlaAlaThrProIleAlaLeu 127
DB 1034 ATTTCAAGACACACAGATGCTTATCTGATTTCTTGAATCGCAAAACCTGCTGAGAT 1093
OY 128 AspValLysLysThrLys-----AspThrLysProValLysLysGlu 142
DB 1094 ACTGTTAAATCAATTGAAATTAGCTGAAGCTAAAGCTTACCTACAGAGAA 1144

RESULT 9
AAN91094
ID AAN91094 standard; DNA: 1950 BP.
XX
AC AAN91094;
XX
DT 04-JUL-1990 (first entry)
XX
DE Protein G gene.
XX
KW Protein G; immunoglobulin; Fc receptor; ds.
XX
OS Streptococcus sp.
XX
FH Key Location/Qualifiers
FT -35_signal 465..470 /+tag- c
FT -10_signal 487..492 /+tag- b
FT RBS 565..571 /+tag- d
FT CDS 578..1918 /*tag- a
FT
FT
XX
PN MO8810306-A.
XX
PD 29-DEC-1988.
XX
PF 20-JUN-1988; 88WO-US02084.
XX
PR 19-JUN-1987; 87US-0063959.
XX
PA (GENE-) GENEX CORP.
XX
PI Fahnestock SR;
XX
DR WPI: 1989-023848/03.
DR P-PSDB; AAP95030.
XX
PT Cloned protein G variant genes -
PT expressing proteins having immunoglobulin-binding properties of
PT protein G and derived from Streptococcus sp.
XX
PS Disclosure; ; 11pp; English.
XX
CC Gene encodes protein G of non-pathogenic streptococcus sp. allowing
CC isolation of the protein and variants, useful as bacterial Fc receptors
CC eg in purification and detection of Abs., screening of hybridoma clones
CC and treatment of disease.
XX
SQ Sequence 1950 BP; 706 A; 324 C; 399 G; 521 T; 0 other;

```

2007-06-01 10:00:00

XX

KW Immunoglobulins; Ig; ds.  
 XX Streptococcus sp GX7809.  
 OS  
 XX Key Location/Qualifiers  
 FH CDS 578..2248  
 FT /tag- a  
 XX  
 XX  
 PN US4977247-A.  
 XX  
 PD 11-DEC-1990.  
 XX  
 XX  
 PF 19-MAY-1989; 89US-0354264.  
 XX  
 PR 19-MAY-1989; 89US-0354264.  
 PR 14-FEB-1986; 86US-0829354.  
 PR 23-APR-1986; 86US-0854887.  
 PR 17-FEB-1987; 87MO-0500329.  
 PR 19-JUN-1987; 87US-0063959.  
 PR 20-JUN-1988; 88US-0209236.  
 XX  
 PA (GENE-) GENEX CORP.  
 XX  
 PI Fahnestoc SR, Lee T, Wroble MH;  
 XX WPI: 1991-006758/01.  
 DR P-PSDB; AAR10004.  
 XX  
 XX  
 PT Immobilised protein G variants - used for detection, isolation  
 PT and purification. Immunoglobulin(s) and immunoglobulin fragments  
 XX  
 PS Disclosure; Fig 8; 52pp; English.  
 XX  
 XX Protein G gene product may be modified allowing the variant to be  
 CC immobilised and exhibit different binding profiles. The bound  
 CC protein is useful in purification and detection of Igs and fragments.  
 XX  
 SO Sequence 1950 BP; 706 A; 323 C; 398 G; 523 T; 0 other;  
 XX  
 Alignment Scores:  
 Pred. No.: 8.12e-11 Length: 1950  
 Score: 188.00 Matches: 61  
 Percent Similarity: 51.59% Conservative: 20  
 Best Local Similarity: 38.85% Mismatches: 60  
 Query Match: 24.20% Indels: 16  
 Gaps: 6  
 US-09-847-539a-6 (1-159) x AAQ10001 (1-1950)  
 QY 1 ValaspeProlleIleProAsnglyThrleuThrasnleu 20  
 DB 677 GTTGATTCACCAATCGAGATACCCCAATTATTCGTAAATGCGTAATTAATCTT 736  
 QY 21 LeuGIysAsnAlaProGIuLyLeuAlaLeuArgAsnGIuGAlaIleAspGIuLeu 40  
 DB 737 CTGGGAAATTCAGAGAACACCTGGCTTGGTAATGAAGAGATGCTACAGCTGATTG 796  
 QY 41 LysLysGlnAlaIleGIuAspLysGIuAlaThrAlaIleGIuAlaIleSerSerasp 60  
 DB 797 ACACGACGAGCGGTAGCGGCTACTGTGGCAGCAGCGAGCTGAATAATGCTGGGCGAGCA 856  
 QY 61 AlaLeuGIuAlaLeuAlaAspGIuThraspAlaLeuInserGIuGIuAlaIleValVal 80  
 DB 857 GCTTGGAGAGCGCGCA--GCAGCAGATGCTCTAGCAAAAGCCAAAGCAGATGCCCTT 913  
 QY 81 LysAlaAspAsn-----AlaAlaSerAspAlaLeuGIuAlaLeu-----AlaAsp 95  
 DB 914 AAAGATTAATCAAAATATGAGTAAGTACATTAACAAGATCTAATCAACAATGCCAAA 973  
 QY 96 GlnThraspAlaLeuGIuInserGIuGIuAlaIleValValGIuInserAspAsnAlaAla 114  
 DB 974 ACTGTGAAGGCAATMAAGACCTTCAAGCACACTGTTGTAATCAAGCAAGAAAGCGCGCT 1033

QY 115 ---SerAspAlaTrpGIu-----LysAlaAlaThrProIleAlaLeu 127  
 DB 1034 ATTTCAGACACACAGATGCGTATCTGATTCTTGAAATCGCAACACCTCGTGAAGAT 1093  
 QY 128 AspValLysLysThrIlys-----AspThrLysProValValLysLysGIu 142  
 DB 1094 ACTGTAAATCAATTAATTAAGCTGAAAGCTAAAGCTTACTACTAAACAGAA 1144  
 RESULT 12  
 AAQ64644  
 ID AAQ64644 standard; DNA: 1950 BP.  
 XX  
 XX  
 AC AAQ64644;  
 XX  
 XX  
 DT 06-JAN-1995 (first entry)  
 XX  
 XX Streptococcus Protein G gene derived from strain GX7809.  
 DE Streptococcus Protein G; variant; IgG binding activity;  
 XX Streptococcus Protein G; variant; IgG binding activity;  
 KM Immunoglobulin; Lancefield Group G; bacterial Fc receptor; ss.  
 XX  
 XX Streptococcus sp. GX7809.  
 OS  
 XX  
 FH Key Location/Qualifiers  
 FH -35\_signal 465..470  
 FT /tag- a  
 FT -10\_signal 487..492  
 FT /tag- b  
 FT RBS 566..572  
 FT /tag- c  
 FT CDS 578..1924  
 FT /tag- d  
 FT /product- Protein\_G  
 XX  
 XX US5312901-A.  
 XX  
 PD 17-MAY-1994.  
 XX  
 XX  
 PF 14-FEB-1986; 86US-0829354.  
 XX  
 PR 14-FEB-1986; 86US-0829354.  
 PR 23-APR-1986; 86US-0854887.  
 PR 19-JUN-1987; 87US-0063959.  
 PR 20-JUN-1988; 88US-0209236.  
 PR 19-JUN-1990; 90US-0540169.  
 PR 21-APR-1992; 92US-0671539.  
 XX  
 XX (PHAA ) PHARMACIA LKB BIOTECHNOLOGY AB.  
 XX  
 XX Fahnestock SR;  
 XX WPI: 1994-159179/19.  
 DR P-PSDB; AAR53290.  
 XX  
 XX New recombinant streptococcal protein G variants - useful for  
 PT antibody detection and purification and for therapy  
 PT  
 PS Example 2; Fig 3; 48pp; English.  
 XX  
 XX A 1.9kb HindIII fragment containing the entire coding sequence for  
 CC Protein G was isolated from Streptococcus GX7809. The protein G has  
 CC IgG-binding activity which has been localised to the B repeating  
 CC structure. Streptococcal Protein G variants comprising the B domains  
 CC are claimed.  
 XX  
 SO Sequence 1950 BP; 706 A; 323 C; 398 G; 523 T; 0 other;  
 XX  
 Alignment Scores:  
 Pred. No.: 8.12e-11 Length: 1950  
 Score: 188.00 Matches: 61  
 Percent Similarity: 51.59% Conservative: 20  
 Best Local Similarity: 38.85% Mismatches: 60  
 Query Match: 24.20% Indels: 16





```

US-09-847-539A-6 (1-159) x AA075036 (1-2383)
OY 1 ValAspSerProIIleGlunInProArGIIleIleProAsnGlyGlyThrIeuThrsAnIeu 20
DB 676 GTTGACTACACCAATCGAAGATACCCCAATTAATTCGTAATGCTGTGTAATTAATTAATCTTT 735
OY 21 LeuGlyAsnAlaProGluLysLeuAlaLeuArGAsnGluGluArGAlaIleAspGluLeu 40
DB 736 CTGGGAATTCAGACAGAACACACTGCTTGGCTAATGAAGAGAGTGTACAGCTGATTTG 795
OY 41 LysLysGlnAlaIleGluAspLysGlnAlaIleThrAlaIleGluAlaIleSerSerAsp 60
DB 796 ACAGCAGACGCGGTACCGCTACTGTGGCAGCAGCGCGCACTGAAATCTGTGGCGCAGCA 855
OY 61 AlaLeuGlnAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaIleVal 80
DB 856 GCTTGGGAAGCAGCGCA---GCAGCAGATGCTCTGACAAAGCCAAAGCAGATGCCCTT 912
OY 81 LysAlaAspAsn-----AlaAlaSerAspAlaLeuGluAlaLeu-----AlaAsp 95
DB 913 AAGGAATTCACAAATATGAGTAGTACTTACAGATCTATCAACATGCAATGCCAA 972
OY 96 GlnThrAspAlaLeuGlnSerGluGlnAlaGluValAlaGlnSerAspAlaIle--- 114
DB 973 ACTGTTGAAGGCGCTAAAGACCTTCAAGCAGCAAGTTGTAATCAGCAGCAAGGCGCT 1032
OY 115 ---SerAspAlaLeuGlnSerGluGlnAlaGluValAlaGlnSerAspAlaIle 127
DB 1033 ATTTCAGAAAGCAACAGATGCTTATCTGATTTCTTAAATTCACAAACCTGCTGAAAGT 1092
OY 128 AspValLysLysThrLys-----AspThrLysProValValLysLysGlu 142
DB 1093 ACTGTTAAATCAATGATTAATGATTAAGCTAAAGCTTAAAGCTTAAAGCAGAGA 1143
RESULT 16
AA247931
ID AA247931 standard; cDNA; 1576 BP.
XX
AC AA247931;
XX
DT 10-MAR-2000 (first entry)
XX
DE Streptococcus strain G 148 protein encoding cDNA.
XX
KW Streptococcus strain G 148; protein G'; protein G primer; liposome;
KW liposomal delivery complex; connecting moiety; antibody; Fc region;
KW diagnostic; therapeutic agent; antibiotic; antidiabetic; antiviral;
KW antitumorigenic; cytokine; hormone; imaging agent; neurotransmitter;
KW stimulant; cytotoxic agent; malignant cell; ss.
XX
OS Streptococcus sp.
XX
FH Key Location/Qualifiers
FT CDS 292..603
FT /*tag= a
FT CDS 705..1262
FT /*tag= b
FT /product= "protein G'"
FT /transl_except= (pos:705..707,aa:Met)
XX
PN W09959643-A2.
XX
PD 25-NOV-1999.
XX
PF 19-MAY-1999; 99WO-US11177.
XX
PR 20-MAY-1998; 98US-0086347.
XX
PA (SDGS-) SDG INC.
XX
PI Lau JR:
XX

```

```

DR WPI: 2000-062383/05.
DR P-PSDB: AAY57610, AAY57611.
XX
PT New Liposome constructs comprising a liposome connected to an antibody,
XX used, e.g. for delivery of cytotoxic agents to malignant cells
XX
PS Disclosure: Fig 1; 32pp: English.
XX
XX
The present invention describes a construct for connecting an antibody
or antibody fragment to a liposome. The construct comprises protein G'
or a linking moiety for connecting the protein G' to the liposome.
CC Also described are: (1) a liposomal delivery vehicle comprising: (a) a
CC liposome; and (b) a connecting moiety connected to the liposome, which
CC specifically binds the Fc region of an antibody, for connecting the
CC antibody to the liposome; and (2) forming a liposomal carrier comprising:
CC (a) providing liposomal components having binding moieties, for forming
CC a liposome; (b) providing a construct comprising a linking moiety and a
CC connecting moiety bound together; and (c) combining the liposomal
CC components and the construct and sonicating the combination so that the
CC binding moieties are exposed on a surface of the liposome, for binding to
CC the linking moiety, and to facilitate the binding between them. The
CC products can be used for the delivery of diagnostic or therapeutic
CC agents. The liposomes may contain or may be associated with a diagnostic
CC or therapeutic agent, e.g. antibiotics, antidiabetics, imaging agents,
CC antitumorigenic, antivirals, cytokines, hormones, neurotransmitters,
CC neurotransmitters, or stimulants. They can be used particularly for the
CC delivery of cytotoxic agents to malignant cells. The protein G'
CC connecting moiety provides a liposomal delivery complex having improved
CC targeting efficiency. As a result of the binding between protein G' and
CC the Fc region of antibodies, protein G' shields the Fc regions of the
CC attached antibodies from non-specific binding to cell-surfaces, other
CC proteins, and anatomical structures. The present sequence encodes
CC an unidentified protein and protein G'.
XX
SQ Sequence 1576 BP; 572 A; 270 C; 330 G; 404 T; 0 other:
Alignment Scores:
Pred. No.: 2,21e-07 Length: 1576
Score: 155.00 Matches: 81
Percent Similarity: 25.75% Conservative: 13
Best Local Similarity: 22.19% Mismatches: 62
Query Match: 19.95% Indels: 210
DB: Gaps: 7
US-09-847-539A-6 (1-159) x AA247931 (1-1576)
OY 1 ValAspSerProIIleGlunInProArGIIleIleProAsnGlyGlyThrIeuThrsAnIeu 20
DB 391 GTTGACTACACCAATCGAAGATACCCCAATTAATTCGTAATGCTGTGTAATTAATTAATCTTT 450
OY 21 LeuGlyAsnAlaProGluLysLeuAlaLeuArGAsnGluGluArGAlaIleAspGluLeu 40
DB 451 CTGGGAATTCAGACAGAACACACTGCTTGGCTAATGAAGAGAGTGTACAGCTGATTTG 510
OY 41 LysLysGlnAlaIleGluAspLysGlnAlaIleThrAlaIleGluAlaIleSerSerAsp 60
DB 511 ACAGCAGACGCGGTACCGCTACTGTGGCAGCAGCGCGCACTGAAATCTGTGGCGCAGCA- 569
OY 61 AlaLeuGlnAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaIleVal 80
DB 570 GCTTGGGAAGCAGCGCA---GCAGCAGATGCTCTGACAAAGCCAAAGCAGATGCCCTT 626
OY 81 LysAlaAspAsnAlaLeuAlaSerAspAlaLeuGlnAlaLeuAlaAspGlnThrAspAlaLeu 100
DB 627 AAGGAATTCACATA-----GATGAATTTTAAGTGTATTAAGTGAAGCTGTGATGCT 680
OY 101 Gln----- 101
DB 681 AATTAATTCCTTAATGTTAAACATTGAAGCGCAAAACAACTAGTAAGCTGTGATGCT 740
OY 101 ----- 101
DB 741 GCTTACTGCAAAAAGCTTCAACAAATAGGTTAAGCAGACGCTGTGACGCTGAATGG 800

```



```

QY 102 -----SerGluGluAlaGluValGln--- 109
Db 801 ACTTACGACGATGCGACTAAGACCTTACAGTTAGTGAACCAAGAGTATGCG 860
QY 109 ----- 109
Db 861 TCTGATTACACGAGCGGTGACAACTTACAACCTTGTATTATGTTAAACATTGAA 920
QY 109 ----- 109
Db 921 GGCAGAACACTACTGAGCTGTGATGCTGCTACTGAGAAAAAGTTCACAACTATC 980
QY 110 ----SeraspAlaAlaSerAspAlaTrp----- 118
Db 981 GCTAACGACAGCGGTGTGACGCTGAATGAGCTTACGACGATGGACTAAGACCTTACA 1040
QY 118 ----- 118
Db 1041 GTTACTGAAAAACGAAAGTATGATGCGTCTGAATTACACAGCGGTGACAACTTAC 1100
QY 118 ----- 118
Db 1101 AACCTGTATTATGTTAAACATTGAAAGCGAACAACACTACTAAAGCAGTAGCGCA 1160
QY 119 ----GluLysAla----- 121
Db 1161 GAAACGACGAAAAAGCCTTCAACAACTACGCTAACGACAGCGTGTGATGCTTTGG 1220
QY 121 ----- 121
Db 1221 ACTTATGATGATGGAGCTAAGACCTTTACGCTAATAAGTTACAGAGTTCCTGG 1280
QY 121 ----- 121
Db 1281 TGATGCACCACTGACCAAGAAAAACGAAAGCAAGTATCCCTGTTCCGTTAACTCC 1340
QY 122 -AlaThrProIleAlaLeuAspValLysLysThrLysAspThr----- 135
Db 1341 TGCACCTCAATTGCTTAAGATGACCGCTAAGAAAGCAGATCTAAGAAAGAGATGCTAA 1400
QY 136 -LysProValValLysLysGluGluArgGlnAsnValAsnThrLeuProThrThrGlyGly 155
Db 1401 AAACCAAGACCTAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1460
QY 155 uGluSerAsnPro 159
Db 1461 AGGAAGCAACCA 1473

RESULT 17
AAN91099 standard; DNA; 798 BP.
AC AAN91099;
XX 04-JUL-1990 (first entry)
DE 'Protein G gene variant.
XX Protein G; Immunoglobulin; Fc receptor; ds.
XX OS Streptococcus sp.
XX PN W08810306-A.
XX 29-DEC-1988.
XX 20-JUN-1988; 88WO-US02084.
XX 19-JUN-1987; 87US-0063959.
XX (GENE-) GENEX CORP.
XX

```

```

PI Fahnestock SR;
XX WPI; 1989-023848/03.
DR P-PSDB; AAP94785.
XX
XX Cloned protein G variant genes -
PT expressing proteins having immunoglobulin-binding properties of
PT protein G and derived from Streptococcus sp.
XX
XX Claim 4; Page 83; 116pp; English.
XX
XX Gene encodes protein G variant of non-pathogenic streptococcus allowing
CC isolation of the protein and variants, useful as bacterial Fc receptors
CC eg in purification and detection of Abs., screening of hybridoma clones
CC and treatment of disease.
XX
SQ Sequence 798 BP; 277 A; 161 C; 177 G; 183 T; 0 other;

Alignment Scores:
Pred. No.: 5 06e-06 Length: 798
Score: 139.00 Matches: 58
Percent Similarity: 35.65% Conservative: 24
Best Local Similarity: 25.22% Mismatches: 64
Query Match: 17.89% Indels: 84
DB: 10 Gaps: 8

US-09-847-539a-6 (1-159) x AAN91099 (1-798)
QY 4 ProIleGluGlnProArgIleIleProAsnGlyLysThrLeuThrAsnLeuGlyAsn 23
Db 100 CCAATCGAGATATACCCCAATTATTCGTAATGCTGTAATTAATTAATCTCTGGGAA 159
QY 24 AlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaIleAspGlu----- 39
Db 160 TCAGAGACACACCTGGCTTGGCTAATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 219
QY 40 -----LeuLysGluGlnAlaIleGluAspLysGluAlaThr 51
Db 220 AAGACGACACTTCAAAATTATTCCTTAATGTGTTAAACATTGAAAGCGAACAACACTACT 279
QY 52 ThrAlaIleGluAlaIleAspSerAsp-----Ala 61
Db 280 GAACCTGTGATGCTGCTACTGACGAGAAAAAGTCTCAACAACTACGCTAACGACACGCT 339
QY 62 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGln-----SerGluGluAla 77
Db 340 GTTGACGCTGAATGAGCTTACGACGATCCGCTAAGACCTTTACAGTTACTGAAAAACCA 399
QY 78 AlaValValLysAlaAsp-----Asn 84
Db 400 GAAGTGCATCGATGCGTCTGAATTAAACACCGCCGTGACAACTTCAAACTTGTATTAA 459
QY 85 AlaIleSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGlu 104
Db 460 GGTAAACATTGAAGAGCGAACAACACTTAACAGACAGTAGAGCAAGAAACGAGAAAAA 519
QY 105 Ala--GluValValGlnSerAspAsnAlaIleSerAspAlaTrp----- 118
Db 520 GCCTCAACAACTACGCTAAGACGACAGCGTGTGATGCTGTGGACTTATGATGATGCG 579
QY 118 ----- 118
Db 580 ACTAAGACCTTTACGCTAAGTGAATGTTACAGAGGTTCTCGAGCTGATGACCAACT 639
QY 119 ----GluLys-----AlaAlaThrProIle 125
Db 640 GAACGAGAAAAACGAAAGCAAGTATCCCTTGTTCCTTAATCTCTCGCAACTCCAACT 699
QY 126 AlaLeuAspValLysLysThrLysAspThr-----LysProValVal 139
Db 700 GCTAAAGATGACGCTAAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGCT 759
QY 140 LysLysGluGluArgGlnAsnValAsnThr 149

```

```

Db      760  AACAAAGATGACGCTAAGAAAGCTGAACCT 789
      |||||:||||: |||
RESULT 18
ID      AAQ06009 standard; DNA; 798 BP.
XX
XX      AAQ06009;
AC
XX      17-JUN-1991 (first entry)
DT
XX      Sequence encoding Protein G variant.
DE
XX      Immunoglobulin; ds.
XX
OS      Streptococcus sp. Lancefield Group G strain.
XX
XX      US4956296-A.
PN
XX      11-SEP-1990.
PD
XX      20-JUN-1988; 88US-0209236.
PF
XX      20-JUN-1988; 88US-0209236.
PR      14-FEB-1986; 86US-0829354.
PR      23-APR-1986; 86US-0854887.
PR      17-FEB-1987; 87MO-US00329.
PR      19-JUN-1987; 87US-0063959.
XX
XX      (GENE-) GENEX CORP.
XX
XX      Fahnestock SR;
PI
XX      WPI: 1990-297491/39.
DR      P-PSDB; AAR07004.
XX
XX      Recombinant Protein G variants - obtd. using a cloned gene
PT      encoding Protein G from Streptococcus sp., used for blinding
PT      immunoglobulin.
XX
XX      Claim 4; Column 47; 48pp; English.
PS
XX
CC      Sequence may be incorporated into a non-pathogenic host eg. E.coli,
CC      where they may be expressed at high levels. The proteins have a
CC      higher binding efficiency and capacity for immunoglobulin, and may
CC      be used for purifying, detecting and isolating antibodies.
XX
SO      Sequence 798 BP; 277 A; 161 C; 177 G; 183 T; 0 other;

Alignment Scores:
Pred. No.:      5,06e-06      Length:      798
Score:          139.00      Matches:      58
Percent Similarity: 35.65%      Conservative: 24
Best Local Similarity: 25.22%      Mismatches: 64
Query Match:    17.89%      Indels:      84
DB:            11      Gaps:      8

US-09-847-539A-6 (1-159) x AAQ06009 (1-798)
OY      4  ProtiEGluGlnProArgIleIleProAsnGlyGlyThrLeuThrAsnLeuGluAsn 23
      |||||||
DB      100  CCAATCGAAGATACCCCAATTATTCGTAAAGTGCGTAATTAATTAATCTTCTGGGGAAT 159
      |||||||
OY      24  AlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaIleAspGlu----- 39
      |||
DB      160  TCAGAGACAACACTGGCTTGGCTAATGAAGAGAGTGCTACAGCTGGTAACCATTAACCT 219
      |||||
OY      40  -----LeuLysIysGlnAlaIleGluAspLysGluAlaThr 51
      |||
DB      220  AAGACTGACACTTAACAATTATTCCTTAATGTAAACAATGTAAAGCGCAAACTACT 279
      |||
OY      52  ThrAlaIleGluAlaAlaSerSerAsp-----Ala 61
      |||

```

```

Db      280  GAAGCTGTTGATGCTGCTACTGACGAGAAAAAGTCTCAACAAATACGCTTAACGACAACGCT 339
OY      62  LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGln-----SerGluAla 77
      |||||
DB      340  GTTGACGCGTGAATGACCTTACGACGATGCGACTTAAGACCTTACAGTTACTGAAAAACA 399
      |||||
OY      78  AlaValValLysAlaAsp-----Asn 84
      |||
DB      400  GAAGTCATCGATCGCTGCTGAATTAACACCAAGCCGTGACAACTTACAACTTGTATTAAAT 459
      |||||
OY      85  AlaAlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGlu 104
      |||
DB      460  GGTAAACAACTGAAGGCAACCACTAAAGCAGTAGACGCAAGAACTGCAGAAAAA 519
      |||
OY      105  Ala---GluValValInserAspAsnAlaAlaSerAspAlaThr----- 118
      |||
DB      520  GCCTTCAACCAATACGCTTAACGACCAAGGTTGATGCTTGGACTTATGATGATGCG 579
      |||
OY      118  ----- 118
DB      580  ACTAAGACCTTTAGCGTTAACTGAATGTTACAGAGTTCCTCGAGTGTATGACCAACT 639
OY      119  -----GluLys-----AlaAlaThrProIle 125
      |||||
DB      640  GAACCAAGAAAAACGACAGCAAGATATCCCTTGTTCGTTAACTCTGCACTCCAACTT 699
      |||||
OY      126  AlaLeuAspValLysLysThrLysAspThr-----LysProValVal 139
      |||
DB      700  GCTTAAGATGACGCTTAAGAAAGACGATACTTAAGAAAGAAAGATGCTTAAAAACCAAGACT 759
      |||||
OY      140  LysLysGluGluArgGlnAsnValAsnThr 149
      |||||
DB      760  AAGAAAGATGACGCTAAGAAAGCTGAACCT 789
      |||||
RESULT 19
ID      AAQ10007 standard; DNA; 798 BP.
XX
XX      AAQ10007;
AC
XX      13-MAR-1991 (first entry)
DT
XX
DE      Type 4 GX7809 protein G variant.
XX
XX      Immunoglobulins; Ig; ds.
XX
XX      Streptococcus sp GX7809.
OS
XX      US497247-A.
PN
XX      11-DEC-1990.
PD
XX      19-MAY-1989; 89US-0354264.
PF
XX      19-MAY-1989; 89US-0354264.
PR      19-MAY-1989; 89US-0354264.
PR      14-FEB-1986; 86US-0829354.
PR      23-APR-1986; 86US-0854887.
PR      17-FEB-1987; 87MO-US00329.
PR      19-JUN-1987; 87US-0063959.
PR      20-JUN-1988; 88US-0209236.
XX
XX      (GENE-) GENEX CORP.
XX
XX      Fahnestock SR, Lee T, Wroble MH;
PI
XX      WPI: 1991-006758/01.
DR      P-PSDB; AAR10011.
XX
XX      Immobilised protein G variants - used for detection, isolation
PT      and purification. Immunoglobulin(s) and immunoglobulin fragments
XX
XX      Disclosure; Column 11; 52pp; English.
XX

```

CC Protein G variant product carries active regions B1 and B2. It may  
 CC be immobilised and exhibits different binding profiles. The bound  
 CC protein is useful in purification and detection of Igs and fragments.

SQ Sequence 798 BP; 277 A; 161 C; 177 G; 183 T; 0 other;

#### Alignment Scores:

Pred. No.:	Length:	Matches:	Score:
5,06e-06	798	58	139.00
Percent Similarity:		Conservative:	24
Best Local Similarity:		Mismatches:	64
Query Match:		Indels:	84
		Gaps:	8

US-09-847-539a-6 (1-159) x AA064648 (1-798)

```

QY 4 ProilegluglnProarglleleProasnlglyThrleuthrasnleuLeuGlyAsn 23
DB 100 CCATCGAAGATACCCCAATTAATTCGTAATGCTGTAATTAATTAATTAATTAAT 159
QY 24 AlaProglulysleuAlaLeuArgAsnGluArgAlaIleAspGlu----- 39
DB 160 TCAGAGACACACCTGCTTGCCTAATGACAGAGAGTGTACAGCTGGGTACCATTA 219
QY 40 -----leuLysylsGlnAlaIleGluAspLysGluAlaThr 51
DB 220 AAGACTGACACTTCAAAATTAATTCCTTAATGTTAAACATTGAAAGCGAACAAC 279
QY 52 ThrAlaIleGluAlaAlaSerSerAsp-----Ala 61
DB 280 GAAGCTGTGATGCTGCTACTGACGAGAAAGTCTTCAACAAATACGCTAACGACA 339
QY 62 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGln-----SerGluGluAla 77
DB 340 GTTACAGCTGTAAGTACGCTTACGACGATGCGACTAAGCTTACAGTTACTGAA 399
QY 78 AlaValIleLysAlaAsp-----Asn 84
DB 400 GAATGTGATCGATGCTGCTGAATTAACACACGACCGCTGACAACTTCAAACT 459
QY 85 AlaAlaSerAspAlaLeuAlaLeuAlaLeuAlaAspGlnThrAspAlaLeuGln 104
DB 460 GCTAAACATGTAAGGCGAAGCAACTAATGACAGTACGACGAGCAAGCAAGCA 519
QY 105 Ala---GluValIleGlnSerAspAlaAlaSerAspAlaTrp----- 118
DB 520 GCCTTCAACAACTACGCTAAGCAGACAGCGTGTGATGCTGATGATGATGAT 579
QY 118 ----- 118
DB 580 ACTAAGACCTTTACGCTAATGTAAGTAAGTGTACAGAGCTTCTCGAGGTGATG 639
QY 119 -----AlaAlaThrProIle 125
DB 640 GAACGAGAAACCAAGCAAGTATCCCTTGTTCGTTAACTCCGCAACTCCAA 699
QY 126 AlaLeuAspValLysLysThrLysAspThr-----LysProValVal 139
DB 700 GCTAAAGATGACGCTAAGCAAGTATGTAAGCAAGATGCTAAGCAAGCAAG 759
QY 140 LysLysGluGluArgGlnAsnValAsnThr 149
DB 760 AAGAAAGATGACGCTAAGCAAGTGAAGCAAGTGAAGCAAGT 789

```

RESULT 20  
 AA064648 standard; DNA; 798 BP.  
 AC AA064648;  
 DT 06-JAN-1995 (first entry)  
 XX Igc-binding streptococcus Protein G variant coding sequence.

```

XX KW Streptococcus Protein G; variant; Igc binding activity;
XX KM Immunoglobulin; Lancefield Group G; bacterial Fc receptor; ss.
XX OS Streptococcus sp.
XX FH Key Location/Qualifiers
XX FT sig-peptide 1..90
XX FT mat-peptide /note="encodes 30 amino acid secretion sequence"
XX FT /tag= b
XX FT /product= Protein_G_variant
XX PN US5312901-A.
XX PD 17-MAY-1994.
XX PF 14-FEB-1986; 86US-0829354.
XX PR 14-FEB-1986; 86US-0829354.
XX PR 23-APR-1986; 86US-0854887.
XX PR 19-JUN-1987; 87US-0063959.
XX PR 20-JUN-1988; 88US-0209236.
XX PR 19-JUN-1990; 90US-0540169.
XX PR 21-APR-1992; 92US-0871539.
XX PA (PHAA ) PHARMACIA LKB BIOTECHNOLOGY AB.
XX PI Fahnstock SR;
XX DR WPI: 1994-159179/19.
XX DR P-PSDB; AAR53294.
XX PT New recombinant streptococcal protein G variants - useful for
XX PT antibody detection and purification and for therapy
XX PS Disclosure; Columns 9-12 : 48pp; English.
XX CC Protein G isolated from Streptococcus has Igc-binding activity which
XX CC has been localised to the B repeating structure (see AAR53290).
XX CC The nucleotide sequence AA064648 codes for a claimed streptococcal
XX CC Protein G variant comprising the B domain active site and retaining
XX CC Igc-binding activity.
XX SQ Sequence 798 BP; 277 A; 161 C; 177 G; 183 T; 0 other;

```

Alignment Scores:

Pred. No.:	Length:	Matches:	Score:
5,06e-06	798	58	139.00
Percent Similarity:		Conservative:	24
Best Local Similarity:		Mismatches:	64
Query Match:		Indels:	84
		Gaps:	8

US-09-847-539a-6 (1-159) x AA064648 (1-798)

```

QY 4 ProilegluglnProarglleleProasnlglyThrleuthrasnleuLeuGlyAsn 23
DB 100 CCATCGAAGATACCCCAATTAATTCGTAATGCTGTAATTAATTAATTAATTAAT 159
QY 24 AlaProglulysleuAlaLeuArgAsnGluArgAlaIleAspGlu----- 39
DB 160 TCAGAGACACACCTGCTTGCCTAATGACAGAGAGTGTACAGCTGGGTACCATTA 219
QY 40 -----leuLysylsGlnAlaIleGluAspLysGluAlaThr 51
DB 220 AAGACTGACACTTCAAAATTAATTCCTTAATGTTAAACATTGAAAGCGAACAAC 279
QY 52 ThrAlaIleGluAlaAlaSerSerAsp-----Ala 61
DB 280 GAAGCTGTGATGCTGCTACTGACGAGAAAGTCTTCAACAAATACGCTAACGACA 339

```

PR	06-MAY-1999	9905-01332486
PR	06-MAY-1999	9905-01332687
PR	07-MAY-1999	9905-01332863
PR	11-MAY-1999	9905-01334256
PR	11-MAY-1999	9905-01342421
PR	14-MAY-1999	9905-01343419
PR	14-MAY-1999	9905-01343720
PR	14-MAY-1999	9905-01343770
PR	18-MAY-1999	9905-01347461
PR	19-MAY-1999	9905-01349418
PR	20-MAY-1999	9905-01351524
PR	21-MAY-1999	9905-01351553
PR	24-MAY-1999	9905-01356529
PR	25-MAY-1999	9905-01366021
PR	27-MAY-1999	9905-01363692
PR	28-MAY-1999	9905-01367812
PR	01-JUN-1999	9905-01372722
PR	03-JUN-1999	9905-01375728
PR	04-JUN-1999	9905-01375702
PR	07-JUN-1999	9905-01377274
PR	08-JUN-1999	9905-01380694
PR	10-JUN-1999	9905-01385409
PR	10-JUN-1999	9905-01388147
PR	14-JUN-1999	9905-01391119
PR	16-JUN-1999	9905-01394552
PR	16-JUN-1999	9905-01394593
PR	17-JUN-1999	9905-01394922
PR	18-JUN-1999	9905-01394542
PR	18-JUN-1999	9905-01394554
PR	18-JUN-1999	9905-01394655
PR	18-JUN-1999	9905-01394617
PR	18-JUN-1999	9905-01394612
PR	18-JUN-1999	9905-01394663
PR	18-JUN-1999	9905-01397570
PR	18-JUN-1999	9905-01397633
PR	21-JUN-1999	9905-01398617
PR	22-JUN-1999	9905-01398617
PR	23-JUN-1999	9905-01403553
PR	23-JUN-1999	9905-01405554
PR	24-JUN-1999	9905-01405655
PR	28-JUN-1999	9905-01409823
PR	29-JUN-1999	9905-01409821
PR	30-JUN-1999	9905-0141287
PR	01-JUL-1999	9905-01418442
PR	01-JUL-1999	9905-01421554
PR	02-JUL-1999	9905-01420555
PR	06-JUL-1999	9905-01423909
PR	08-JUL-1999	9905-01428053
PR	09-JUL-1999	9905-0142820
PR	12-JUL-1999	9905-01428277
PR	13-JUL-1999	9905-01435344
PR	14-JUL-1999	9905-01435324
PR	15-JUL-1999	9905-01440055
PR	16-JUL-1999	9905-01440055
PR	16-JUL-1999	9905-01440865
PR	19-JUL-1999	9905-01443331
PR	19-JUL-1999	9905-01443332
PR	19-JUL-1999	9905-01443333
PR	19-JUL-1999	9905-01443334
PR	19-JUL-1999	9905-01443335
PR	20-JUL-1999	9905-01443552
PR	20-JUL-1999	9905-01446352
PR	20-JUL-1999	9905-01446884
PR	21-JUL-1999	9905-01446114
PR	21-JUL-1999	9905-01450866
PR	21-JUL-1999	9905-01450888
PR	22-JUL-1999	9905-01450875
PR	22-JUL-1999	9905-01450877





PF 06-SEP-1994: 94WO-SE00826.  
 XX  
 PR 06-SEP-1993: 93SE-0002855.  
 XX  
 PA (GUSS/) GUSS B.  
 PA (JONS/) JONSSON H.  
 PA (LIND/) LINDBERG M.  
 PA (MUEL/) MUELLER H.  
 PA (RANT/) RANTAMAKI L K.  
 XX  
 PI Guss B, Jonsson H, Lindberg M, Mueller H, Rantamaki LK;  
 XX  
 DR WPI: 1995-123382/16.  
 DR P-PSDB; AAR71928.  
 XX  
 PT DNA encoding fast alpha 2-macroglobulin-binding proteins - used  
 PT to obtain prods. for sepn., detection or quantification or for  
 PT binding inhibition  
 XX  
 PS Disclosure; Fig.4; 50pp; English.  
 XX  
 CC A phage lambda EMBL library of S. dysgalactiae 8215 DNA was analyzed  
 CC for fast-acting alpha-2-macroglobulin (FAM)-, BSA- and IgG-binding  
 CC activity. Clone lambda-SD1, which expressed all 3 activities, was  
 CC analyzed to obtain DNA encoding the FAM-binding protein, MAG. The  
 CC mag gene (given in Q89197) encodes a 44 kDa protein (R71929).  
 XX  
 SQ Sequence 1555 BP; 583 A; 242 C; 304 G; 426 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 0.00489 Length: 1555  
 Score: 114.50 Matches: 49  
 Percent Similarity: 34.228 Conservative: 15  
 Best Local Similarity: 26.208 Mismatches: 56  
 Query Match: 14.74% Indels: 67  
 DB: 16 Gaps: 6  
 US-09-847-539a-6 (1-159) x AAQ89197 (1-1555)  
 QY 36 AAlaIleaspGluLeuLysLysGlnAlaIleGluAsp----- 47  
 DB 903 GCTATTGAAATTCCTGAGAAATTCGCCGCTTACTATTAAATTAAATTAAT 962  
 QY 48 -----LysGluAlaTrpThrAlaIleGlu-----AlaIleSerSerAspAla 61  
 DB 963 GGTAAACTGCAGAGCTGTGACTGCTCTTAAGATGAATTTAGCTCAAAACAGCA 1022  
 QY 62 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValLys 81  
 DB 1023 GTGATTGACGACCTGAATTAACACGACCTTGACAACTCAAACTTGTATC----- 1076  
 QY 82 AlaAspAsnAlaAlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGln 101  
 DB 1077 -----AATGGTAAACACTGAAAGCGCAAAACACTACTAAAGCAGTAGACGAGAAACT 1130  
 QY 102 SerGlnGluAla---GluValValGlnSerAspAsnAlaAlaSerAspAlaTrpGlu--- 119  
 DB 1131 GCAGAAAAAGCCTTCAACAACATACGCTAACGAAAAAGCGTGTGATGCTTTGGACTTAC 1190  
 QY 119 ----- 119  
 DB 1191 GATGATCGACCTAGACCTTACTGTAAGTGAATGCTTACTGAAGTTCCTGGATGCA 1250  
 QY 120 -----LysAlaAlaThr 123  
 DB 1251 CCAACTGACCAAAAAACGAGAACAGATCCCTCTTGTTCGTTAACTCCTGCAACT 1310  
 QY 124 ProIleAlaLeuAspValLysThrLysAspThr----- 135  
 DB 1311 CCAATTGCTTAAGATGACGCTTAAGAAAGACGATCTAAGAAAGCAGATCTAAGAAAGAA 1370  
 QY 136 -----LysProValValLysLysGlnGluArgGlnAsnValAsnThrLeuProThr 152  
 PA 136 -----LysProValValLysLysGlnGluArgGlnAsnValAsnThrLeuProThr 152

DB 1371 GATGCTAAAAAACGAGACTAGACAAAGAAAGAACTAAGAAAGCTGCACCTTCTTACA 1430  
 QY 153 ThrGlyGluGlnSerAsnPro 159  
 DB 1431 ACTGGTAGAGAAACCAACCA 1451  
 RESULT 24  
 AAN70757  
 ID AAN70757 standard; DNA; 1469 BP.  
 XX  
 AC AAN70757;  
 XX  
 DT 21-MAY-1991 (first entry)  
 XX  
 DE Sequence encoding polypeptide possessing IgG-binding activity of  
 DE protein G from Streptococcus G148.  
 XX  
 KW Antibody-binding; IgG; IgA; immunoglobulin; ss.  
 XX  
 OS Streptococcus G148.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 1..1443  
 FT /\*tag- a  
 FT 9..80  
 FT /\*tag- b  
 FT /label- A1  
 FT 81..233  
 FT /\*tag- c  
 FT /label- B1  
 FT 234..305  
 FT /\*tag- d  
 FT /label- A2  
 FT 306..458  
 FT /\*tag- e  
 FT /label- B2  
 FT 459..530  
 FT /\*tag- f  
 FT /label- A3  
 FT 531..566  
 FT /\*tag- g  
 FT /label- S  
 FT 567..731  
 FT /\*tag- h  
 FT /label- C1  
 FT 732..776  
 FT /\*tag- i  
 FT /label- D1  
 FT 777..941  
 FT /\*tag- j  
 FT /label- C2  
 FT 942..986  
 FT /\*tag- k  
 FT /label- D2  
 FT 987..1151  
 FT /\*tag- l  
 FT /label- C3  
 FT 1152..1361  
 FT /\*tag- m  
 FT /label- W  
 FT 1362..1469  
 FT /\*tag- n  
 FT /label- M  
 XX  
 PN W08705631-A.  
 PD 24-SEP-1987.  
 XX  
 PF 20-MAR-1987; 87WO-SE00145.  
 XX  
 PR 21-MAR-1986; 86SE-0001325.  
 XX  
 PA (PHAA ) PHARMACIA AB.

PA (GUS5/) GUS5 B M.  
XX  
XX Guss BM, Lindberg KM, Flock JI, Uhlen CEM;  
XX WPI: 1987-277686/39.  
XX P-PSDB: AAF70468.  
XX  
XX New recombinant DNA molecules - for producing proteins with  
XX IgG-binding specificity of protein G or proteins A and G  
XX  
XX Example: Fig 2; 39pp: English.  
XX  
XX A recombinant DNA molecule containing a nucleotide sequence which  
XX codes for a protein or polypeptide having the same IgG specificity as  
XX protein G from *Streptococcus* G148 (AAN70757) is claimed. See, for  
XX example, AAN70754, AAN70755 and AAN70756.  
XX  
XX Sequence 1469 BP; 545 A; 280 C; 297 G; 347 T; 0 other;  
SQ  
Alignment Scores:  
Pred. No.: 0.00847 Length: 1469  
Score: 112.00 Matches: 44  
Percent Similarity: 39.44% Conservative: 12  
Best Local Similarity: 30.99% Mismatches: 62  
Query Match: 14.41% Indels: 24  
DB: Gaps: 5  
US-09-847-539A-6 (1-159) x AAN70757 (1-1469)  
QY 41 LysGlnAlaIleGluAspLysGluAlaThrAlaIleGlu-----Ala 56  
DB 943 AACCAGAAAGTATCGATCGTGTGAATTACACACCGCGTCAACTTCAAACTTGT 1002  
QY 57 AlaSerSerAspAlaLeu-----GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSer 74  
DB 1003 ATTAAATGGTAAACATTGAAAGCCGAAACACTACTTAAAGCAGTACGAGAACTGCA 1062  
QY 75 GluGluAlaIleValValLysAlaAspAlaIleSerAspAlaLeuGluAlaLeuAla 94  
DB 1063 GAAAAAGCCTTCAACAAATACGCTAACGACAAAGCGTGTGTGGACTTATGAT 1122  
QY 95 AspGlnThrAspAlaLeuGlnSerGluGluAlaIleValValGlnSerAspAlaIle 114  
DB 1123 GATGGCAGTAAAGACCTTACGCTTACGTAATG---GTTACAGAGGTTCTCGTGTATGCA 1179  
QY 115 SerAspAlaThrGluLys-----AlaAlaThr 123  
DB 1180 CCAACTGAACCAAGAAAAACAGACAGATATCCCTCTTCCGTTAACTCCTGCAACT 1239  
QY 124 ProIleAlaLeuAspValLysLysThrLysAspThr-----LysPro 137  
DB 1240 CCAATTGCTAAAGATGACGCTAAAGAAAGACGATACAGAAAGAGATGCTAAAAACCA 1299  
QY 138 ValValLysLysGluGluAlaIleGlnSerValAsnThrLeuProThrThrGluGlnSer 157  
DB 1300 GAAGCTTAAAGAAAGAGACGCTAAAGAAAGCTGAAGAACTTCTCAACAGTGTGTAAGGACGC 1359  
QY 158 AsnPro 159  
DB 1360 AACCCA 1365  
RESULT 25  
ID ABL20337 standard: DNA; 2121 BP.  
XX ABL20337;  
XX ABL20337;  
XX 26-MAR-2002 (first entry)  
XX Drosophila melanogaster genomic polynucleotide seq ID NO 12484.  
XX Drosophila: developmental biology; cell signalling; insecticide;  
XX pharmaceutical; gene; ds.

XX OS Drosophila melanogaster.  
XX  
XX PN WO200171042-A2.  
XX  
XX 27-SEP-2001.  
XX  
XX 23-MAR-2001; 2001WO-US09231.  
XX  
XX 23-MAR-2000; 2000US-191637P.  
XX  
XX 11-JUL-2000; 2000US-0614150.  
XX  
XX (PEKE ) PE CORP NY.  
XX  
XX Venter JC, Adams M, Li PMD, Myers EW;  
XX WPI: 2001-656860/75.  
XX  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
XX genes from *Drosophila* and for elucidating cell signalling and cell-cell  
XX interactions -  
XX  
XX Claim 1: SEQ ID NO 12484; 21pp + Sequence Listing; English.  
XX  
XX The invention relates to an isolated nucleic acid detection reagent  
XX capable of detecting 1000 or more genes from *Drosophila*. The invention is  
XX useful in developmental biology and in elucidating cell signalling and  
XX cell-cell interactions in higher eukaryotes for the development of  
XX insecticides, therapeutics and pharmaceutical drugs. The invention  
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
XX sequences (ABB57737-ABB72072).  
XX  
XX The sequence data for this patent did not form part of the printed  
XX specification, but was obtained in electronic format directly from WIPO  
XX at ftp.wipo.int/pub/published\_pcc\_sequences.  
XX  
XX Sequence 2121 BP; 533 A; 638 C; 587 G; 363 T; 0 other;  
SQ  
Alignment Scores:  
Pred. No.: 0.0169 Length: 2121  
Score: 111.00 Matches: 41  
Percent Similarity: 46.92% Conservative: 20  
Best Local Similarity: 31.54% Mismatches: 43  
Query Match: 14.29% Indels: 26  
DB: Gaps: 6  
US-09-847-539A-6 (1-159) x ABL20337 (1-2121)  
QY 4 ProIleGluGlnProAlaGlnIleProAsnGlyGlyThrLeuThrAsnLeuGlnLysAsn 23  
DB 1150 CCCCTGAGCAGATCCCTCGATGTG-----GAG 1176  
QY 24 AlaProGluLysLeuAlaLeuArgAsnGlu-----GluArgAlaIle 37  
DB 1177 GCCAAGGAAACAAACGCTGCGCAGAGCGCTTCTTACCGTTTACCAAGAGCGCTTC 1236  
QY 38 AspGluLeuLysLysGlnAlaIleGluAspLysGluAlaIleThrThrAlaIleGluAlaIle 57  
DB 1237 GACGAGGTGAC-----TTGCAAAAGAACCGTCCAGAACCCATCAAGCAG 1287  
QY 58 SerSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeu-----GlnSerGlu 75  
DB 1288 CGTAGGAAAGAG---AGGCTCTCAAGAGAGAGAGATCGCTGACGCTGAGCCGAA 1344  
QY 76 GluAlaIleValValLysAlaAspAlaIleSerAspAlaLeuGluAlaLeuAlaAsp 95  
DB 1345 AAGCAGCGTGTCTGAAGCAAAAAGCTGCGAGAGGCGCCCAAAATCGCTGCCGA 1404  
QY 96 Gln-----ThrAspAlaLeuGlnSerGluGluAlaIleValValGlnSerAspAsn 112  
DB 1405 GAGGCTCTCTGCTCAAGCTGCCCAAAAAGCAGCGCAAGAGCTTAAAGCCCTTAAA 1464  
QY 113 AlaAlaSerAspAlaThrPgiLysAlaIle 122



Db 1465 GCCGCTGAAGATGCTGCCCAAGAAAGCTGCC 1494  
|||||  
RESULT 26  
AAAX13365/c  
ID AAAX13365 standard; DNA: 3212 BP.  
XX  
AC AAAX13365;  
XX  
DT 19-MAR-1999 (first entry)  
XX  
DE Enterococcus faecalis genome contig SEQ ID NO:428.  
XX  
KM Enterococcus faecalis; contig: detection; Enterococcal infection;  
KW vaccine; attenuation; computer readable medium; ds.  
XX  
OS Enterococcus faecalis.  
XX  
PN MO9850555-A2.  
XX  
PD 12-NOV-1998.  
XX  
PE 04-MAY-1998; 98WO-US08985.  
XX  
PR 14-NOV-1997; 97US-0066009.  
PR 06-MAY-1997; 97US-0044031.  
PR 16-MAY-1997; 97US-0046655.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Barash SC, Dillon PJ, Kunsch CA;  
XX  
XX WPI; 1999-045171/04.  
XX  
PT New isolated Enterococcus faecalis polynucleotides and polypeptides  
PT - used to develop products for the detection of Enterococcus and for  
PT use in vaccines for prevention or attenuation of Enterococcus  
PT infection.  
XX  
PS Claim 1; Page 1650-1652; 2084pp; English.  
XX  
XX A computer readable medium has been developed which has recorded on it  
CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.  
CC AAAX12938 to AAAX13919 represent these nucleotide sequences which are  
CC primary nucleotide sequences, also known as contigs. The computer-based  
CC system can identify fragments of the Enterococcus faecalis genome with  
CC commercial importance. The products can be used to detect the presence  
CC of Enterococcus faecalis in samples. They can also be used for  
CC diagnosing Enterococcal infection in an animal and monitoring  
CC progression of disease, and for identifying agents which can be used to  
CC modulate the growth or pathogenicity of Enterococcus faecalis, or  
CC another related organism, in vivo or in vitro. In particular the  
CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences  
CC can be used in vaccines to prevent or attenuate an Enterococcal  
CC infection.  
XX  
XX Sequence 3212 BP: 843 A; 664 C; 504 G; 1182 T; 19 other;  
SO  
Alignment Scores:  
Pred. No.: 0.0359 Length: 3212  
Score: 110.00 Matches: 44  
Percent Similarity: 45.03% Conservative: 24  
Best Local Similarity: 29.14% Mismatches: 71  
Query Match: 14.16% Indels: 12  
DB: 20 Gaps: 6  
US-09-847-539A-6 (1-159) x AAAX13365 (1-3212)  
Oy 7 GlnProAgtIleIleProAsnGlclyGlyThrLeuThrAsnLeuLeuGlyAsnAlaProGlu 26  
Db 663 CAACCAAAACACCTGAAACACGTTCTTCA-----GAACCAACCAACA 622  
Oy 27 LysLeuAlaLeuArgAsnGluArgAlaIleAspGluLeuLysGluAlaIleGlu 46

Db 621 GTGAAGCTACCAACAAACAGCAGCAGCTATTACTGA---AAACAGCAACAGTAA 565  
Oy 47 AspLysGluAlaThrThrAlaIleGluAlaAlaSerSerAspAlaLeuGluAlaLeuAla 66  
Db 564 GAGAAACACAGCAATTGTCGATCAAAACCAACAGTCCGCTGACACTGCGGAAAAAGAAAA 505  
Oy 67 AspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValLysAlaAspAsnAlaAla 86  
Db 504 GACGCCATTGATCAATCTGTTAAAGACCAACACACCTGCTGATCAAAACAAAGACCA 445  
Oy 87 SerAsp--AlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluAlaAla 105  
Db 444 TTGATCAAGATCAACAAACAGCAGTACGACCA---CAAGCAGTTGTAAGCAAGTAA 388  
Oy 106 GluValValGlnSerAspAsnAlaAlaSerAspAlaTrpGluLysAlaIleThrProIle 125  
Db 387 AAGTCTGTG-----GATGAAGCAACACCTTCACCATTTGAAAAAGCCAAAGACCAAGTG 334  
Oy 126 AlaLeuAspValLysLysThrLysAspThrLysProValLysLysGluArgGln 145  
Db 333 GCTACGCTACACAGCGCGTTGATGACCAACAAAGTAGTR---GATCAAGCTCAACA 277  
Oy 146 AsnValAsnThrLeuProThrThrGlyGluGlu 156  
Db 276 GACGTTAACCAACAAACAGCTGTGTCGAAGAA 244  
RESULT 27  
AAC38790  
ID AAC38790 standard; DNA: 812 BP.  
XX  
AC AAC38790;  
XX  
XX 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 22241.  
XX  
KW Hybridisation assay; genetic mapping; gene expression control;  
KW protein identification; signal transduction pathway;  
KW metabolic pathway; promoter; termination sequence; ss.  
XX  
OS Arabidopsis thaliana.  
XX  
XX EP1033405-A2.  
XX  
XX 06-SEP-2000.  
XX  
XX 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.

```
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139753.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144332.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.

PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 04-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 09-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149930.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151338.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 18-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
```

PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161922.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

## Alignment Scores:

Pred. No.:	0.00765	Length:	812
Score:	109.50	Matches:	41
Percent Similarity:	43.68%	Conservative:	35
Best Local Similarity:	23.56%	Mismatches:	65
Query Match:	14.09%	Indels:	33
DB:	21	Gaps:	6

US-09-847-539A-6 (1-159) x AAC38790 (1-812)

QY 7 GlnProArglle-----lIleProAsnGlyGlyThrLeuThrAsnLeu 20

Db 87 CACCCCAAAATCACAAAAAATAATATATATATGCGTGC----- 131

QY 21 LeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluArgAlaIle----- 37

Db 132 TGTCGAGTAGACCAAGGAATCATACATGCCAAACGAGAGGCTGTTCTTACAG 191

QY 38 -----AspGluLeuLysLys 42

Db 192 CCTGCTCTGAAAAAGTAGTAGCCAGGAGAACACAGAAAGCGGTGACAAACAGAAC 251

QY 43 GlnAlaIleGluAspLysGluAlaThrThralaIleGluAlaIleSerSerAspAlaLeu 62

Db 252 CAGACAGTGGCTGAGACACGTGAACAACTCTGCGAAGCTAAGAGACATCTCCGCTC 311

QY 63 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluAlaIleAlaValLysAla 82

Db 312 GAGCCGACCAAGAAACACTCTGCTGACACCGAGGTGACCGCTGAGAGAGTCT 371

QY 83 AspAsnAlaIleSerAspAlaLeuGluAlaLeuAla--AspGlnThrAspAlaLeuGln 101

Db 372 TCTTCTGCTGCTGCTGAGGAGCGACGACATTGCACTGAGAAAGTGAGAAAGCTGCA 431

QY 102 SerGluGluAlaGluAlaValGlnSerAspAsnAlaIleSerAspAlaIleTrpGluLysAla 121

Db 432 ACAGAAAAAGCGCTGAA--GCAGAAAGGTGAGGCGGCTGACGTCACCCGAAAG-- 485

QY 122 AlaThrProIleAlaLeuAspValLysThrLysAspThrLysProValValLysLys 141

Db 486 ---CTTGAGTTCGACGAGCGGAGCGGAGGAAAGCGCTGAGAGTGAACGTAAGGTTGAG 542

QY 142 GluGluArgGlnAsnValAsnThrLeuProThrThrGlyGlu 155

Db 543 GCTGACCAAGTAAAGCGTGAAGCTGACCAAGTAAAGCTGAG 584

RESULT 28 AAC46811 ID AAC46811 standard; DNA; 1128 BP.

XX AAC46811;

XX AC 18-OCT-2000 (first entry)

XX Arabidopsis thaliana DNA fragment SEQ ID NO: 51490.

XX Hybridisation assay; genetic mapping; gene expression control;

XX protein identification; signal transduction pathway;

XX metabolic pathway; promoter; termination sequence; ss.

XX Arabidopsis thaliana.

PN	EP1033405-A2.
XX	
PD	06-SEP-2000.
XX	
PF	25-FEB-2000; 2000EP-0301439.
XX	
PR	25-FEB-1999; 99US-0121825.
PR	05-MAR-1999; 99US-0123180.
PR	09-MAR-1999; 99US-0123548.
PR	23-MAR-1999; 99US-0125788.
PR	25-MAR-1999; 99US-0126264.
PR	29-MAR-1999; 99US-0126785.
PR	01-APR-1999; 99US-0127462.
PR	06-APR-1999; 99US-0128234.
PR	08-APR-1999; 99US-0128714.
PR	16-APR-1999; 99US-0129845.
PR	19-APR-1999; 99US-0130077.
PR	21-APR-1999; 99US-0130449.
PR	23-APR-1999; 99US-0130510.
PR	28-APR-1999; 99US-0130891.
PR	30-APR-1999; 99US-0132048.
PR	03-APR-1999; 99US-0132407.
PR	04-MAY-1999; 99US-0132484.
PR	05-MAY-1999; 99US-0132485.
PR	06-MAY-1999; 99US-0132486.
PR	07-MAY-1999; 99US-0132487.
PR	11-MAY-1999; 99US-0134256.
PR	14-MAY-1999; 99US-0134218.
PR	14-MAY-1999; 99US-0134219.
PR	14-MAY-1999; 99US-0134221.
PR	14-MAY-1999; 99US-0134370.
PR	18-MAY-1999; 99US-0134768.
PR	19-MAY-1999; 99US-0134941.
PR	20-MAY-1999; 99US-0135124.
PR	21-MAY-1999; 99US-0135353.
PR	24-MAY-1999; 99US-0135629.
PR	25-MAY-1999; 99US-0136021.
PR	27-MAY-1999; 99US-0136352.
PR	28-MAY-1999; 99US-0136782.
PR	01-JUN-1999; 99US-0137222.
PR	03-JUN-1999; 99US-0137528.
PR	04-JUN-1999; 99US-0137502.
PR	07-JUN-1999; 99US-0137724.
PR	08-JUN-1999; 99US-0138094.
PR	10-JUN-1999; 99US-0138540.
PR	10-JUN-1999; 99US-0138847.
PR	14-JUN-1999; 99US-0139119.
PR	16-JUN-1999; 99US-0139452.
PR	16-JUN-1999; 99US-0139453.
PR	17-JUN-1999; 99US-0139452.
PR	18-JUN-1999; 99US-0139454.
PR	18-JUN-1999; 99US-0139455.
PR	18-JUN-1999; 99US-0139456.
PR	18-JUN-1999; 99US-0139457.
PR	18-JUN-1999; 99US-0139458.
PR	18-JUN-1999; 99US-0139459.
PR	18-JUN-1999; 99US-0139460.
PR	18-JUN-1999; 99US-0139461.
PR	18-JUN-1999; 99US-0139462.
PR	18-JUN-1999; 99US-0139463.
PR	18-JUN-1999; 99US-0139750.
PR	18-JUN-1999; 99US-0139763.
PR	21-JUN-1999; 99US-0139817.
PR	22-JUN-1999; 99US-0139839.
PR	23-JUN-1999; 99US-0140353.
PR	23-JUN-1999; 99US-0140354.
PR	24-JUN-1999; 99US-0140635.
PR	28-JUN-1999; 99US-0140823.
PR	29-JUN-1999; 99US-0140991.
PR	30-JUN-1999; 99US-0141287.
PR	01-JUL-1999; 99US-0141842.

```
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 23-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.

PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 26-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Alignment Scores:
Pred. No.: 0.0129
Score: 109.00
Percent Similarity: 39.20%
Best Local Similarity: 27.27%
Query Match: 14.03%
DB: 21
Gaps: 7

US-09-847-539a-6 (1-159) x AAC46811 (1-1128)
QY 27 LysLeuAlaLeuArgAsnGluIuArgAlaLeuAspGluLeuLys----- 42
DB 571 AAGCTCGACCTTCANAAACGTGCGAAGACACATTTCAGACCTTAACGATTGAGCAAGA 630
QY 43 -----GlnAlaIleGluAsp-----LysGluAlaThrThra 53
DB 631 GAGAGCGATCATCAACCATTTGAGATTCGAGAAAGCAAAAGATGCAAAACC 690
QY 54 ILeGluAlaLaserSerSpAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGln 73
DB 691 CTCCTGCTGTAGAACAGACGTCAGATTGCTGTTTGAAGCTCAGAACGAAAGAT 750
QY 74 SerGluGlu----- 76
DB 751 GCAGAGCAACACGCCGAGGACCGCTAAATGTTCAGATTTCGATTCAGATCA 810
QY 77 AlaAlaValAlaLysAlaAspAsnAla-----AlaSerAspAlaLeuGluAla 92
DB 811 TTACGCTGTGTGAAGAAAGTGCTGTGTAAGATTACAGACAGACACACTTAAAGAGCT 870
QY 93 LeuAlaAspGlnThr---AspAlaLeuGlnSerGluAlaGluValGlnSerAsp 111
```

DB 871 TGCAGATGAAACACTCAACATTAGAGATGAGAGAGAAATGAGATATATATAA 930  
 QY 112 AsnAlaIaIaSerAspAlaTtPgluLysAla-----AlaThrProIleAlaLeu 127  
 DB 931 GCTGCACACACAGATCTTTGAAGAAAGCTGAGATGGCTCACAAGCTACAAATTTGGTG 990  
 QY 128 AspValLysLys----- 131  
 DB 991 GAGCTTGAGCTGAAGAGCGCGTAAAGCAGCAGTAGATCTTTGGCAGATCCAAATG 1050  
 QY 132 ---ThLysAspThrLysProValValLysLysGluGluArgGlnAsn 146  
 DB 1051 TGTGCCAAATCTACCAAGAGAGTGTCTAGTCAAAACCAAGATCATCG 1098  
 RESULT 29  
 ABL11003  
 ID ABL11003 standard; cDNA; 4086 BP.  
 AC ABL11003;  
 XX  
 DT 26-MAR-2002 (first entry)  
 DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 27491.  
 XX  
 KM Drosophila; developmental biology; cell signalling; insecticide;  
 KM pharmaceutical; gene; ss.  
 XX  
 OS Drosophila melanogaster.  
 XX  
 PN WO200171042-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 23-MAR-2001; 2001WO-US09231.  
 XX  
 PR 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 XX  
 PA (PEKE ) PE CORP NY.  
 XX  
 PI Venter JC, Adams M, Li PWD, Myers EW;  
 XX  
 DR MPI; 2001-656860/75.  
 DR P-PSDB; ABB66900.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 XX  
 PS Claim 1: SEQ ID NO 27491; 21bp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins  
 CC (AB857737-AB872072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SO Sequence 4086 BP; 1311 A; 960 C; 1056 G; 759 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 0.0789 Length: 4086  
 Score: 108.00 Matches: 36  
 Percent Similarity: 47.27% Conservative: 42  
 Best Local Similarity: 21.82% Mismatches: 61  
 Query Match: 13.90% Indels: 26  
 DB: 23 Gaps: 6

US-09-847-539a-6 (1-159) x ABL11003 (1-4086)  
 QY 11 IleProAsnGlyGlyThrLeuThrAsnLeuLeuGlyAsnAlaProGluLysLeuAlaLeu 30  
 DB 2796 GTCCCAAT-----ATTAATGTCAGAGCCGAGAAATGATGTG 2834  
 QY 31 ArgAsnGlu-----GluArgAlaIleAspGluLeuLysLysGlnAla 44  
 DB 2835 GACTCACAAACGGCGGACGAGAGAGCTTCGACAGCCGACGATCCAGATCAAGAAAGAGCT 2894  
 QY 45 IleGluAspLysGluAlaThrThrAlaIleGluAlaIaSer-----SerAsp 60  
 DB 2895 GACCTTGAAACATCAAGAGAGATTTATCAAGAGATTCACCAATTCGCATCCGAT 2954  
 QY 61 AlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaValAla 80  
 DB 2955 TCCGTGGCGAATTAAGAGAGCGGTAACTGCTCTGAGGGCGAGATGATGTCATCTA 3014  
 QY 81 LysAlaAspAsnAlaIaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeu 100  
 DB 3015 GAGCCGATACATACGCCAAAGAGCTCTGATGATTAATGCGGAGCCGAGAAA--- 3071  
 QY 101 GlnSerGluGluAlaGluValAlaGlnSer---AspAsnAlaIaSerAspAlaTtPglu 119  
 DB 3072 CCTGATCAGAGAGAGACATGTTCAATCTGAAGAGAAATCGACAGACAGAGCGTTAGAT 3131  
 QY 120 LysAlaIaIaThr-----ProIleAlaLeuAspValLysLysThr 132  
 DB 3132 AGGTCACTAACGACGAGAGATGACCTGTGCCACCTACGAGGTACTTACCAACAATG 3191  
 QY 133 LysAspThrLysProValValLysLysGluGluArgGlnAsnValAsnThrLeuProThr 152  
 DB 3192 GAGATTGATGACACGACGCGGAGAAAGCAGCCGAAATATATGACTTACAAACGCA 3251  
 QY 153 ThrGlyGluLys 157  
 DB 3252 GACGAGAAAGAGCC 3266  
 RESULT 30  
 ABL11002/c  
 ID ABL11002 standard; cDNA; 6782 BP.  
 XX  
 AC ABL11002;  
 XX  
 DT 26-MAR-2002 (first entry)  
 DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 27488.  
 XX  
 KM Drosophila; developmental biology; cell signalling; insecticide;  
 KM pharmaceutical; gene; ss.  
 XX  
 OS Drosophila melanogaster.  
 XX  
 PN WO200171042-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 23-MAR-2001; 2001WO-US09231.  
 XX  
 PR 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 XX  
 PA (PEKE ) PE CORP NY.  
 XX  
 PI Venter JC, Adams M, Li PWD, Myers EW;  
 XX  
 DR MPI; 2001-656860/75.  
 DR P-PSDB; ABB66899.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 XX



```

QY 124 ProIleAlaLeuAspValLysThrLysAspThrLysProValLysGluGln 143
Db 190 CAAGTGGCTACGATACACAGCTGTGATGACCAACAAAAGTAGTCAGTCAAGCTCAG 131
QY 144 ArgGlnAsnValAsn 148
Db 130 CAGACGTTAACCAAC 116

RESULT 32
AAT91293
ID AAT91293 standard; DNA; 989 BP.
XX
AC AAT91293;
XX
DT 27-APR-1998 (first entry)
XX
DE Rabbit enteropathogenic E. coli (RDEC-1) espB gene.
XX
KW EspB; rabbit enteropathogenic E. coli; RDEC-1; virulence factor;
KW pathogen; infection; diagnosis; therapy; vaccine; gastroenteritis;
KW ss.
XX
OS Escherichia coli.
XX
FH Key Location/Qualifiers
FT RBS 19..21
FT CDS /*tag= a
FT 31..975
FT /*tag= b
XX
PN MO9740063-A2.
XX
PD 30-OCT-1997.
XX
PF 23-APR-1997; 97MO-CA00265.
XX
PR 23-APR-1996; 96US-0015999.
XX
PA (UYBR-) UNIV BRITISH COLUMBIA.
XX
PI Finlay BB, Kenny B, Stein M;
XX
DR WPI: 1997-535772/49.
DR P-PSDB: AAN27286.
XX
PT EspA from entero-pathogenic or entero-haemorrhagic E. coli - used to
PT immunise against, and treat diseases caused by EspA producing
PT organisms
XX
PS Example 6; Fig 4B; 62pp; English.
XX
CC This sequence comprises the espB gene from a rabbit
CC enteropathogenic Escherichia coli (RDEC-1) strain. It codes for a
CC claimed EspB virulence factor (see AAN27286). The espA and espA
CC (see AAT91292) genes were obtained by PCR from RDEC-1 chromosomal
CC DNA using primers derived from the published sequence of
CC enteropathogenic E. coli. Claimed novel, isolated EspA polypeptides
CC are characterised by: (a) being a secreted protein from
CC enteropathogenic (EPEC) or enterohaemorrhagic (EHEC) E. coli;
CC and (b) having a mol. wt. of about 25 kD as determined by SDS-PAGE.
CC Isolated espA nucleic acids can be used in the recombinant
CC production of EspA polypeptides, and as probes in the diagnosis of
CC disease caused by EspA-producing E. coli. A claimed method of
CC producing a non-pathogenic E. coli comprising inserting a
CC selectable marker nucleic acid into the espA gene. EspA
CC polypeptides can be used to immunise a host susceptible to disease
CC caused by EspA-producing E. coli, and in a method for ameliorating
CC disease caused by EspA-producing organisms.
XX
SQ Sequence 989 BP; 278 A; 211 C; 240 G; 260 T; 0 other;

```

Alignment Scores:

```

Pred. No.: 0.0262 Length: 989
Score: 105.50 Matches: 45
Percent Similarity: 44.858 Conservative: 29
Best Local Similarity: 27.276 Mismatches: 72
Query Match: 13.588 Indels: 20
DB: 18 Gaps: 6

US-09-847-539A-6 (1-159) x AAT91293 (1-989)
QY 4 ProIleGluGlnProArgIleIleProAsnGlyGlyThrLeuThrAsnLeuGlyAsn 23
Db 305 CCGCTGCACACGCCGT-CTGGTT-----GTTGGGCTATTTCATCAGTATGGGGATC 357
QY 24 AlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaIleAspGluLeuLysGln 43
Db 358 TTAGGCTCTTTTGCAGCAATTAACAGTACGCAAAAGCCGAGTGAT----- 405
QY 44 AlaIleGluAspLysGluAlaThrThr-----AlaIleGluAlaIleAspSerAsp 60
Db 406 ---ATTGCTCAAAAACCCGCTCTACATCTTTAGGCTATTTGATGGCTTCTGATACT 462
QY 61 AlaLeuGluAlaLeuAlaAspGlnThr-----AspAlaLeuGlnSerGluGln 76
Db 463 GCGACTAAACGTTGACTAGGCAACGCAAGCGTTGCTGATGCTGTTGAAGATGCATCC 522
QY 77 AlaAlaValAlaLysAlaAspAsnAlaIleSerAspAlaLeuGluAlaLeuAlaAspGln 96
Db 523 AGCGTGTGCGAGCAGGAGTACGACTACGACGACGAGGCGCCGCTACATCCGAGCT 582
QY 97 ThrAspAlaLeuGlnSerGluGluAlaGluValValGlnSerAspAsnAlaIleSerAsp 116
Db 583 GGTGATGACATTCGCCATTTTGTCTCAGACAGCTTCTAGCTGCTGCAAAACGCTGCAGAT 642
QY 117 AlaTrpGluLysAlaIleThrPro-----IleAlaLeuAspValLysLysThr 132
Db 643 GCCGTCAGAGGCAAGTCCGCGCAAGCCGCTTATGCTGCAGTAGAT---AAGATTACT 699
QY 133 LysAspThrLysProValLysLysGluGluArgGlnAsnValAsnThrLeuProThr 152
Db 700 GGCTTACACCATTAATGCGCTTACCAGTCTTGCCGAGCAGAGACATATGGCCACA 759
QY 153 ThrGlyGluGlnSer 157
Db 760 ACCGTTATCTGATCA 774

RESULT 33
AA084886
ID AA084886 standard; cDNA; 627 BP.
XX
AC AA084886;
XX
DT 11-OCT-1995 (first entry)
XX
DE Synthetic protein G gene, MD1.
XX
KW Synthetic; protein G; IgG; Fc receptors; heavy chain; leukaemia;
KW lymphoma; cancer; autoimmune disease; ds.
XX
OS Synthetic.
XX
PN WO9506125-A.
XX
PD 02-MAR-1995.
XX
PF 23-AUG-1994; 94MO-US09141.
XX
PR 23-AUG-1993; 93US-0110653.
XX
PA (IMMU-) APPLIED IMMUNE SCI INC.
XX
PI Lee YM, Okarma TB, Talib S;
XX
SQ WPI: 1995-106854/14.

```





Db 1832 ACTTTCGA-----GGCGAACAACACTACTAAAGAGTA-----GAGCAGAAACT 1876  
QY 36 AlaIleaspgIuLeuLysLysGlnAlaIleGluasp----- 47  
Db 1877 GCACAAAAAGCCTTCAACACTACGTACGAAAAAGCGGTGTTACGCGTAAGTCTTAT 1936  
QY 48 LysGlnIaIaThrThrAlaIleGlnAlaIaIaSeraspAlaIaGlnAlaIaLeuAlaasp 67  
Db 1937 GAGCATGCACTAAACCTTTACAGTTCAGTAAACCAACCGAGTGATGACGCACTGAA 1996  
QY 68 GlnThraspAlaIaLeuGlnInserGluGlnAlaIaIaValIaIaLysAlaAspAlaIaIaSer 87  
Db 1997 TTACACACAGCATTGACAACTACAACTGTATC-----AATGGTAAACA 2044  
QY 88 AspAlaIaLeuGlnAlaIaLeuAlaIaAspGlnThraspAlaIaLeuGlnInserGluGlnAlaIa---Glu 106  
Db 2045 TTGAAGGCGCAACACTACTAAAGCAGTACGCCAGAACTGCAAAAAAGCCTTCAAA 2104  
QY 107 ValValGlnInserAspAsnAlaIaIaIaSeraspAlaIaTrp----- 118  
Db 2105 CAATACGCTAACGAAACGGTGTTGATGCTTTGGACTTACGATGATGCGACTAACGAC 2164  
QY 119 -----GluLys 120  
Db 2165 TTACGGTAAGTAAATGGTACTGAAAGTTCCTGGTATGCAACCACTGAAACGAAAG 2224  
QY 121 -----AlaIaIaTrpProIleAlaIaLeuAspAla 129  
Db 2225 CCAGAGCAGATATCCCTCTGTTCCGTTAACTCCTGCACACTCCAAATGCTAAAGATGAC 2284  
QY 130 LysLysThrLysAspThrLysProVal----- 138  
Db 2285 GCTAAGAAAGACGATCTACTAAGAAAGTCGATCTAAGAAAGAGACGCTAAACCAACAGAA 2344  
QY 139 -----ValLysLysGlnGluIaIaIaGlnAsnValaIaIaTrpLeuProThrThr 153  
Db 2345 GCTAAGAAAGAGAGAGCTAAGAAAGAGAGAGCTAAGAAAGCTGCAACTCTTCTACACT 2404  
QY 154 GLyGluGlnSerAsnPro 159  
Db 2405 GGTGAGAGCAACCA 2422  
RESULT 35  
AAC45194  
ID AAC45194 standard; DNA; 575 BP.  
XX  
AC AAC45194;  
XX  
DT 18-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 45636.  
XX  
KW Hybridisation assay; genetic mapping; gene expression control;  
KW protein identification; signal transduction pathway;  
KW metabolic pathway; promoter; termination sequence; ss.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132487.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144088.  
PR 16-JUL-1999; 99US-0144088.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.





PI Yamamoto RT, Xu HH;  
 XX MPI: 2001-611495/70.  
 DR P-PSDB: AAU37373.  
 XX  
 PT New polynucleotides for the identification and development of  
 PT antibiotics, comprise sequences of antisense nucleic acids -  
 XX  
 PS Claim 27; Seq ID No 8869; 511pp; English.  
 XX  
 CC The invention relates to antisense inhibitors of genes essential to  
 CC prokaryotic cellular proliferation, their use in identifying the  
 CC genes, their use in the discovery of novel antibiotics, the essential  
 CC genes themselves and the encoded proteins. The prokaryotes used are  
 CC *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella*  
 CC *pneumoniae*, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The  
 CC invention is also useful for the identification of potential new targets  
 CC for antibiotic development. The antisense nucleic acids can also be used  
 CC to identify proteins used in proliferation, to express these proteins,  
 CC and to obtain antibodies capable of binding to the expressed proteins.  
 CC The proteins can be used to screen compounds in rational drug discovery  
 CC programmes. The antisense nucleic acid sequence is also useful to screen  
 CC for homologous nucleic acids which are required for cell proliferation in  
 CC a wide variety of organisms. The present sequence encodes an  
 CC essential prokaryotic cellular proliferation protein.  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC ftp://ipo.int/pub/published\_pct\_sequences.  
 XX  
 SO Sequence 7437 BP; 3199 A; 1326 C; 1298 G; 1614 T; 0 other:  
 Alignment Scores:  
 Pred. No.: 0.439 Length: 7437  
 Score: 104.00 Matches: 51  
 Percent Similarity: 37.80% Conservativity: 28  
 Best Local Similarity: 24.40% Mismatches: 70  
 Query Match: 13.38% Indels: 60  
 DB: 23 Gaps: 8  
 US-09-847-539a-6 (1-159) x AAS5232 (1-7437)  
 QY 1 ValAspSerProIle---GluGlnProArgIleIleProAsnGlyThrLeuThrAsn 19  
 DB 4130 GTTGATGACAACTGCTGACAACTCAAAATCAAGCAATGATATCAACTGCTGCTACAACT 4209  
 QY 20 LeuLeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGlnArgAla----- 36  
 DB 4210 CAAGACAAAATATCGACAAAAGATTAGTTTAAAGCTTAAAGCAAAAAGCGTTCAAGAT 4269  
 QY 37 -----IleAspGluLeuLysLysGlnAlaIleGluAsp 47  
 DB 4270 ATCTAAATGACAAACAACTAATGATGTTACCAAAATTAAGATCAAGCAGTTGCTGAT 4329  
 QY 48 LysGlu-----AlaThrThrAlaIleGluAlaAlaSerSerAspAlaLeuGluAla 64  
 DB 4330 ATTCAAGTATTACTGACATACAAACATTAAGATGTCGCAAGATGATTAAGCAACA 4389  
 QY 65 LeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaValValLysAlaAspAsn 84  
 DB 4390 AAAGCAACGACAAAAGCGTTATTGCAAACTGCA-----GAT 4431  
 QY 85 AlaAlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSer----- 102  
 DB 4432 GCGACTACTGCAAGAAAAGAAAAGCAAAATCAACAGTAGACCAATTAACAAAGT 4491  
 QY 103 -----GluGluAlaGluValValGlnSerAspAsnAlaAlaSerAspAlaTTP 118  
 DB 4492 AATCAAAATATTGAAATGACAGATCATGATGATTAACACTGCAAAAGATTAATGCA 4551  
 QY 119 GluLysAlaAlaThrProIle-----AlaLeuAspValLys----- 130  
 DB 4552 ATTCAAGCAATTGACCAATCAAGCATCAACGATGTTAAAGCAAGTCAAGACGCGAA 4611

QY 130 ----- 130  
 DB 4612 TTGCTACTGAAATGCAAAATATAATACTGAAATACTTAATATATGAGACTACTAAT 4671  
 QY 131 -----LysThrLysAspThrLysProVal-----ValLysLysGluGlnArgGlnAsn 146  
 DB 4672 GAAGCAAAAAGCTAAGTAAATTGACACAGTAAACAGCATATCAAGCAAGTTTAAATTAAT 4731  
 QY 147 ValAspThrLeuProThrThrGlyGlu 155  
 DB 4732 ATTAATGACCAACTACTACAGGTGAT 4758  
 RESULT 38  
 AAS52178  
 ID AAS52178 standard; DNA: 6228 BP.  
 XX  
 AC AAS52178:  
 XX  
 DT 13-FEB-2002 (first entry)  
 XX  
 DE *Staphylococcus aureus* DNA for cellular proliferation protein #595.  
 XX  
 KW Antisense; ds; prokaryotic cellular proliferation gene;  
 XX antibiotic; antibacterial; drug design.  
 OS *Staphylococcus aureus*.  
 XX  
 PN WO200170955-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 XX 21-MAR-2001; 2001MO-US09180.  
 PF  
 XX 21-MAR-2000; 2000US-191078P.  
 PR 23-MAY-2000; 2000US-206848P.  
 PR 26-MAY-2000; 2000US-207727P.  
 PR 23-OCT-2000; 2000US-242578P.  
 PR 27-NOV-2000; 2000US-253625P.  
 PR 22-DEC-2000; 2000US-257931P.  
 PR 16-FEB-2001; 2001US-269308P.  
 XX  
 PA (ELIT-) ELITRA PHARM INC.  
 XX  
 PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;  
 PI Yamamoto RT, Xu HH;  
 XX  
 DR MPI: 2001-611495/70.  
 DR P-PSDB: AAU34319.  
 XX  
 PT New polynucleotides for the identification and development of  
 PT antibiotics, comprise sequences of antisense nucleic acids -  
 XX  
 PS Claim 27; Seq ID No 4760; 511pp; English.  
 XX  
 CC The invention relates to antisense inhibitors of genes essential to  
 CC prokaryotic cellular proliferation, their use in identifying the  
 CC genes, their use in the discovery of novel antibiotics, the essential  
 CC genes themselves and the encoded proteins. The prokaryotes used are  
 CC *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella*  
 CC *pneumoniae*, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The  
 CC invention is also useful for the identification of potential new targets  
 CC for antibiotic development. The antisense nucleic acids can also be used  
 CC to identify proteins used in proliferation, to express these proteins,  
 CC and to obtain antibodies capable of binding to the expressed proteins.  
 CC The proteins can be used to screen compounds in rational drug discovery  
 CC programmes. The antisense nucleic acid sequence is also useful to screen  
 CC for homologous nucleic acids which are required for cell proliferation in  
 CC a wide variety of organisms. The present sequence encodes an  
 CC essential prokaryotic cellular proliferation protein.  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at





Oy	16	ThleuIthrnsleuleuglysnalaiaProglulysleualaleuArgAsnglucguarg	35
Dd	4000	AGCTTGACACACACAGGTTAAAGAAGCTTAGACAGTTCGAATGGCTTAAGATAAAA--	4056
Oy	36	AlaIleaspcIleuleuylsyglnalaile-----GlusplysglalaIthThrvala	53
Dd	4057	-----CAATTAGGTAGAGAACGAMTCGACMACAAGAAAGAAAGAACCTACTACATAATA	4107
Oy	54	IleglualaIalaseraSeraspalaleuglualeualaaspGIthrasp---Alaleu	72
Dd	4108	GTAACAGAACTTCACAGACGATACGCCCAAGCGATACGAGACACATCCAAACTTGTCAGAT	4167
Oy	73	gInsergluIuaIaIaValaIValyslaIaaspaSnalaIaSeraspalAleugluIa	92
Dd	4168	GAACATTAAGCAGCGCGAAATTTAAAAAGCTRACTAGCGCTGTGCAGAAAACTTTACAGACC	4227
Oy	93	leuulaaspGIthraspalaleugInserglugluIaIgluIaValaIcInSeraspasn	112
Dd	4228	ATTCGCTACCATGACGATTAAG--CGTAGCGAACAAGCGAAAAAGCTCCAAGCC-----	4278
Oy	113	AlaIalaseraSpaIatrpGuIyslaIaIaIthPrOIealaIeuspalyalLylsystr	132
Dd	4279	-----CTAGCGAGATCTTGAAAAAGCT	4299
Oy	133	LysaspThrLyrsProValValLysLygsIgIuIarGlnAsnValaIasnThrLeuProthr	152
Dd	4300	AAGACAAACACAGAAATTCAGATTAAGAGCTGCAGATTGATAGCTGACATACATCTGTGAAA	4359
Oy	153	Thrlglycu	155
Dd	4360	GATGTGTAG	4368

Search completed: October 13, 2002, 05:02:14  
Job time : 240.471 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2002, 03:04:37 : Search time 64.6203 Seconds  
(without alignments)  
273.300 Million cell updates/sec

Title: US-09-847-539A-6

Perfect score: 777

Sequence: 1 VDSPIEQPRIPNGGTLTNL.....KKEERQVNTLPTTGESNP 159

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1980.DAT:\*  
2: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1981.DAT:\*  
3: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1982.DAT:\*  
4: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1983.DAT:\*  
5: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1984.DAT:\*  
6: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1985.DAT:\*  
7: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1986.DAT:\*  
8: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1987.DAT:\*  
9: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1988.DAT:\*  
10: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1989.DAT:\*  
11: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1990.DAT:\*  
12: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1991.DAT:\*  
13: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1992.DAT:\*  
14: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1993.DAT:\*  
15: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1994.DAT:\*  
16: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1995.DAT:\*  
17: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1996.DAT:\*  
18: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1997.DAT:\*  
19: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1998.DAT:\*  
20: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA2000.DAT:\*  
21: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA2001.DAT:\*  
22: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	777	100.0	159	21	AAV71041
2	777	100.0	184	21	AAV71039
3	777	100.0	217	21	AAV71042
4	760	97.8	167	21	AAV71046
5	750	96.5	155	21	AAV71044
6	694	89.3	271	21	AAV71045
7	682	89.1	259	21	AAV71043
8	682	87.8	141	21	AAV71040
9	285	36.7	58	21	AAV71037
10	188	24.2	448	8	AAV70493
11	188	24.2	448	10	AAV95030

12	188	24.2	448	11	AAV7013	Protein G variant
13	188	24.2	448	12	AAV1000	Streptococcus GX78
14	188	24.2	448	15	AAV3290	Streptococcus Prot
15	187	24.1	593	11	AAV7014	Protein G variant
16	187	24.1	593	15	AAV62944	Streptococcus Prot
17	187	24.1	594	12	AAV10005	Streptococcus GX78
18	142	18.3	103	21	AAV57611	Streptococcus stra
19	139	17.9	235	10	AAV94785	Protein G variant.
20	139	17.9	235	11	AAV7004	Protein G variant.
21	139	17.9	265	12	AAV10011	Type 4 GX7809 prot
22	139	17.9	265	15	AAV3294	IGG-binding Strept
23	129	16.6	28	21	AAV71038	Streptococcus pyog
24	120	15.4	23	21	AAV71036	Streptococcus pyog
25	116	14.9	166	21	AAV38475	Streptococcus thall
26	115	14.8	225	21	AAV71048	Streptococcus pyog
27	114.5	14.7	413	16	AAV71928	S. dysgalactiae MA
28	114.5	14.7	413	16	AAV71670	S. dysgalactiae MA
29	112	14.4	154	21	AAV38476	Arabidopsis thall
30	112	14.4	480	8	AAV70468	Sequence of polype
31	109	14.0	191	21	AAV41392	Arabidopsis thall
32	109	14.0	230	21	AAV41381	Arabidopsis thall
33	109	14.0	316	21	AAV41350	Arabidopsis thall
34	108.5	14.0	180	21	AAV20162	Arabidopsis thall
35	108	13.9	1096	22	AAV62318	Drosophila melanog
36	105.5	13.6	168	21	AAV20163	Arabidopsis thall
37	105.5	13.6	314	18	AAV27286	Rabbit enteropatho
38	105	13.5	208	16	AAV71127	Synthetic protein
39	105	13.5	664	16	AAV71929	S. dysgalactiae M1
40	104.5	13.4	151	21	AAV37153	Arabidopsis thall
41	104	13.4	2478	22	AAV43320	Staphylococcus aur
42	104	13.4	2478	22	AAV37374	Staphylococcus aur
43	103.5	13.3	1822	13	AAV27374	Extracellular fact
44	103.5	13.3	2076	22	AAV34319	Staphylococcus aur
45	103.5	13.3	2186	22	AAV37320	Staphylococcus aur

#### ALIGNMENTS

RESULT 1  
AAV71041  
ID AAV71041 standard; peptide: 159 AA.  
XX  
AC AAV71041:  
XX  
DT 29-AUG-2000 (first entry)  
XX  
XX Streptococcus pyogenes strain SF370 GRAB protein fragment #5.  
DE  
XX GRAB protein; protein G related alpha2M binding protein; vaccine;  
KW alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;  
KW Immune response; Streptococcus pyogenes infection.  
XX  
XX Streptococcus pyogenes.  
OS  
PN WO200026240-A2.  
XX  
XX 11-MAY-2000.  
PD  
XX 02-NOV-1999; 99WO-GB03631.  
PF  
XX 02-NOV-1998; 98GB-0023975.  
PR  
XX (ACT-) ACTINOVA LTD.  
PA  
XX Bjorck LH, Rasmussen M;  
PI  
XX WPI; 2000-365572/31.  
DR  
XX New alpha2M binding protein for generating a protective immune response  
PT to group A streptococcus and purifying the binding protein  
PS Claim 5; Page 57-58; 67pp; English.

XX The patent discloses a new family of proteins termed GRAB (protein G  
CC related alpha2M binding protein) from Streptococcus pyogenes which have  
CC the ability to bind alpha2-macroglobulin (alpha2M) and show homology to  
CC protein G of group G Streptococcus. GRAB protein and peptides derived  
CC from it are used in vaccine compositions for generating a protective  
CC immune response against group A Streptococcus. Antibodies against GRAB  
CC are useful for treating Streptococcus pyogenes infections. The protein  
CC is also useful for purifying alpha2M from a sample. The present sequence  
CC is a fragment of GRAB protein from S. pyogenes strain SF370 corresponding  
CC to residues 34-192 and devoid of the membrane spanning region. This  
CC fragment is useful in vaccine composition.  
XX  
SQ Sequence 159 AA:  
  
Query Match 100.0%; Score 777; DB 21; Length 159;  
Best Local Similarity 100.0%; Pred. No. 1.2e-61;  
Matches 159; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 VDSPIDPRLIPNGGTLTNLGNAPKELALRNEERAIDELKQAIEDKETTATIEAASSD 60  
Db 1 VDSPIDPRLIPNGGTLTNLGNAPKELALRNEERAIDELKQAIEDKETTATIEAASSD 60  
  
QY 61 ALEALADQTDALQSEEAAYVKAADNASDALEALADQTDALQSEEAAYVQSDNNAADAMEK 120  
Db 61 ALEALADQTDALQSEEAAYVKAADNASDALEALADQTDALQSEEAAYVQSDNNAADAMEK 120  
  
QY 121 AATPIALDVKKTKDTKRPVKKERQNVNTLPTTGEESNP 159  
Db 121 AATPIALDVKKTKDTKRPVKKERQNVNTLPTTGEESNP 159  
  
RESULT 2  
AA71039  
ID AAY71039 standard; Protein; 184 AA.  
XX  
AC AAY71039;  
XX  
DT 29-AUG-2000 (first entry)  
XX  
DE Streptococcus pyogenes strain SF370 mature GRAB protein.  
XX  
KW GRAB protein; protein G related alpha2M binding protein; vaccine;  
KW alpha2-macroglobulin; group A Streptococcus; GAS; antidiabetic;  
XX Immune response; Streptococcus pyogenes infection.  
XX  
OS Streptococcus pyogenes.  
XX  
FH Key Location/Qualifiers  
FT Binding-site 1..58 /label= alpha2-macroglobulin\_binding-site  
FT Region 59..86 /label= R1  
FT /note= "repeat region"  
FT 67..114 /label= R2  
FT /note= "repeat region"  
FT 115..159 /label= "repeat region"  
FT Region 150..155 /label= Cell\_wall\_spanning\_region  
FT /note= "consensus sequence for gram-positive  
FT surface cell wall anchored proteins"  
FT 160..184 /label= Membrane\_spanning\_region  
FT Region  
PN WO200026240-A2.  
PD 11-MAY-2000.  
XX  
XX 02-NOV-1999; 99WO-GH03631.  
PF 02-NOV-1998; 96GB-0023975.  
XX  
XX

PA (ACTI-) ACTINOVA LTD.  
XX  
XX Bjorck LH, Rasmussen M;  
PI WPI; 2000-365572/21.  
XX  
DR  
XX  
PT New alpha2M binding protein for generating a protective immune response  
to group A streptococcus and purifying the binding protein  
XX  
PS Claim 5; Page 56; 67pp; English.  
XX  
CC The patent discloses a new family of proteins termed GRAB (protein G  
CC related alpha2M binding protein) from Streptococcus pyogenes which have  
CC the ability to bind alpha2-macroglobulin (alpha2M) and show homology to  
CC protein G of group G Streptococcus. GRAB protein and peptides derived  
CC from it are used in vaccine compositions for generating a protective  
CC immune response against group A Streptococcus. Antibodies against GRAB  
CC are useful for treating Streptococcus pyogenes infections. The protein  
CC is also useful for purifying alpha2M from a sample. The present sequence  
CC is a mature GRAB protein from S. pyogenes strain SF370 without the  
CC signal sequence. This sequence is capable of binding alpha2M  
CC and useful in vaccine composition.  
XX  
SQ Sequence 184 AA:  
  
Query Match 100.0%; Score 777; DB 21; Length 184;  
Best Local Similarity 100.0%; Pred. No. 1.4e-61;  
Matches 159; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 VDSPIDPRLIPNGGTLTNLGNAPKELALRNEERAIDELKQAIEDKETTATIEAASSD 60  
Db 1 VDSPIDPRLIPNGGTLTNLGNAPKELALRNEERAIDELKQAIEDKETTATIEAASSD 60  
  
QY 61 ALEALADQTDALQSEEAAYVKAADNASDALEALADQTDALQSEEAAYVQSDNNAADAMEK 120  
Db 61 ALEALADQTDALQSEEAAYVKAADNASDALEALADQTDALQSEEAAYVQSDNNAADAMEK 120  
  
QY 121 AATPIALDVKKTKDTKRPVKKERQNVNTLPTTGEESNP 159  
Db 121 AATPIALDVKKTKDTKRPVKKERQNVNTLPTTGEESNP 159  
  
RESULT 3  
AA71042  
ID AAY71042 standard; Protein; 217 AA.  
XX  
AC AAY71042;  
XX  
DT 29-AUG-2000 (first entry)  
XX  
DE Streptococcus pyogenes strain SF370 full-length GRAB protein.  
XX  
KW GRAB protein; protein G related alpha2M binding protein; vaccine;  
KW alpha2-macroglobulin; group A Streptococcus; GAS; antidiabetic;  
XX Immune response; Streptococcus pyogenes infection.  
XX  
OS Streptococcus pyogenes.  
XX  
FH Key Location/Qualifiers  
FT Peptide 1..33 /label= Signal\_sequence  
FT Protein 34..217 /label= Mature\_GRAB\_protein  
FT Binding-site 34..91 /label= alpha2M\_binding-site  
FT 34..68 /label= "shows homology to E domain of protein G"  
FT 92..119 /label= Repeat\_region\_1  
FT Region 120..147 /label= Repeat\_region\_2  
FT 148..192 /label= Cell\_wall\_spanning\_region  
FT Region  
FT

```

FT Region 183..188
FT /note="consensus sequence for gram-positive
FT surface cell wall anchored proteins"
FT Region 193..217
FT /label="Membrane-spanning-region"
XX
XX WO200026240-A2.
XX
XX 11-MAY-2000.
XX
XX 02-NOV-1999; 99WO-GH03631.
XX
XX 02-NOV-1998; 98GB-0023975.
XX
XX (ACT1-) ACTINOVA LTD.
XX
XX Bjorck LH, Rasmussen M;
XX
XX WPI; 2000-365572/31.
XX
XX N-PSDB; AAD00559, AAD00560.
XX
XX
XX PT New alpha2M binding protein for generating a protective immune response
XX to group A streptococcus and purifying the binding protein
XX
XX PS Claim 5; Fig 2B; 67pp; English.
XX
XX The patent discloses a new family of proteins termed GRAB (protein G
XX related alpha2M binding protein) from Streptococcus pyogenes which have
XX the ability to bind alpha2-macroglobulin (alpha2M) and show homology to
XX protein G of group G Streptococcus. GRAB protein and peptides derived
XX from it are used in vaccine compositions for generating a protective
XX immune response against group A Streptococcus. Antibodies against GRAB
XX are useful for treating Streptococcus pyogenes infections. The protein
XX is also useful for purifying alpha2M from a sample. The present sequence
XX is a full-length GRAB protein from S. pyogenes strain SF370.
XX
SQ Sequence 217 AA:
XX
Query Match 100.0%; Score 777; DB 21; Length 217;
Best Local Similarity 100.0%; Pred. No. 1.7e-61;
Matches 199; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1 VDSPIEQPRITPNGGTITNLGNAPKIALRNERAIDELKKAIEDKEATTATAEASSD 60
Db 34 VDSPIEQPRITPNGGTITNLGNAPKIALRNERAIDELKKAIEDKEATTATAEASSD 93
QY 61 ALALADQTDALQSEEAAYVAVKADNAAADALEALADQTDALQSEEAAYVQSDNAAADAMEK 120
Db 94 ALALADQTDALQSEEAAYVAVKADNAAADALEALADQTDALQSEEAAYVQSDNAAADAMEK 153
QY 121 AATPIALDVKTKDKTPVVKKEERQNVNTLPTTGEE 159
Db 154 AATPIALDVKTKDKTPVVKKEERQNVNTLPTTGEE 192
XX
XX
XX RESULT 4
XX ID AAY71046 standard; Protein; 167 AA.
XX
XX AC AAY71046;
XX
XX DT 29-AUG-2000 (first entry)
XX
XX DE Streptococcus pyogenes strain KTL3 partial GRAB protein.
XX
XX KW GRAB protein; protein G related alpha2M binding protein; vaccine;
XX alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;
XX immune response; Streptococcus pyogenes infection.
XX
XX OS Streptococcus pyogenes.
XX
XX WO200026240-A2.
XX

```

```

PD 11-MAY-2000.
XX
XX 02-NOV-1999; 99WO-GH03631.
XX
XX 02-NOV-1998; 98GB-0023975.
XX
XX (ACT1-) ACTINOVA LTD.
XX
XX Bjorck LH, Rasmussen M;
XX
XX WPI; 2000-365572/31.
XX
XX N-PSDB; AAD00564.
XX
XX PT New alpha2M binding protein for generating a protective immune response
XX to group A streptococcus and purifying the binding protein
XX
XX PS Claim 5; Page 62-63; 67pp; English.
XX
XX The patent discloses a new family of proteins termed GRAB (protein G
XX related alpha2M binding protein) from Streptococcus pyogenes which have
XX the ability to bind alpha2-macroglobulin (alpha2M) and show homology to
XX protein G of group G Streptococcus. GRAB protein and peptides derived
XX from it are used in vaccine compositions for generating a protective
XX immune response against group A Streptococcus. Antibodies against GRAB
XX are useful for treating Streptococcus pyogenes infections. The protein
XX is also useful for purifying alpha2M from a sample. The present sequence
XX is a partial GRAB protein from S. pyogenes strain KTL3.
XX composition.
XX
SQ Sequence 167 AA:
XX
Query Match 97.8%; Score 760; DB 21; Length 167;
Best Local Similarity 100.0%; Pred. No. 4e-60;
Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1 VDSPIEQPRITPNGGTITNLGNAPKIALRNERAIDELKKAIEDKEATTATAEASSD 60
Db 12 VDSPIEQPRITPNGGTITNLGNAPKIALRNERAIDELKKAIEDKEATTATAEASSD 71
QY 61 ALALADQTDALQSEEAAYVAVKADNAAADALEALADQTDALQSEEAAYVQSDNAAADAMEK 120
Db 72 ALALADQTDALQSEEAAYVAVKADNAAADALEALADQTDALQSEEAAYVQSDNAAADAMEK 131
QY 121 AATPIALDVKTKDKTPVVKKEERQNVNTLPTTGEE 156
Db 132 AATPIALDVKTKDKTPVVKKEERQNVNTLPTTGEE 167
XX
XX
XX RESULT 5
XX ID AAY71044 standard; Protein; 155 AA.
XX
XX AC AAY71044;
XX
XX DT 29-AUG-2000 (first entry)
XX
XX DE Streptococcus pyogenes strain AP1 partial GRAB protein.
XX
XX KW GRAB protein; protein G related alpha2M binding protein; vaccine;
XX alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;
XX immune response; Streptococcus pyogenes infection.
XX
XX OS Streptococcus pyogenes.
XX
XX FH Key Location/Qualifiers
XX FT Misc-difference 17
XX FT /note="Thr at position 18 of GRAB protein fragment
XX FT (AAY71036) from S. pyogenes strain SF370 is replaced
XX FT with Ile"
XX
XX PN WO200026240-A2.
XX

```

PD 11-MAY-2000.  
XX  
XX 02-NOV-1999; 99WO-GB03631.  
XX  
XX 02-NOV-1998; 98GB-0023975.  
XX  
XX (ACTI-) ACTINOVA LTD.  
XX  
XX Bjorck LH, Rasmussen M;  
XX  
XX WPI: 2000-365572/31.  
XX  
XX N-PSDB: AAD00562.  
XX  
XX New alpha2m binding protein for generating a protective immune response  
XX  
XX to group A streptococcus and purifying the binding protein -  
XX  
XX Claim 5; Page 60-61; 67pp; English.  
XX  
XX The patent discloses a new family of proteins termed GRAB (protein G  
XX  
XX related alpha2m binding protein) from Streptococcus pyogenes which have  
XX  
XX the ability to bind alpha2-macroglobulin (alpha2m) and show homology to  
XX  
XX protein G of group G Streptococcus. GRAB protein and peptides derived  
XX  
XX from it are used in vaccine compositions for generating a protective  
XX  
XX immune response against group A Streptococcus. Antibodies against GRAB  
XX  
XX are useful for treating Streptococcus pyogenes infections. The protein  
XX  
XX is also useful for purifying alpha2m from a sample. The present sequence  
XX  
XX is a partial GRAB protein from S. pyogenes strain AP1.  
XX  
XX The protein has alpha2m binding region and is useful in vaccine  
XX  
XX composition.  
XX  
XX Sequence 155 AA:  
SQ  
Query Match 96.5%; Score 750; DB 21; Length 155;  
Best Local Similarity 99.4%; Pred. No. 2.9e-59;  
Matches 154; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 2 DSPTEOPRIIPNGGTLTNLGNAPKALRNEERAIDELKKAIEDKATTAIEAASSDA 61  
DB 1 DSPTEOPRIIPNGGTLTNLGNAPKALRNEERAIDELKKAIEDKATTAIEAASSDA 60  
QY 62 LBALDQDTALQSEEAAYVKADNAASDALEALADQDTALQSEEAAYVSDNNAASDAWEKA 121  
DB 61 LBALDQDTALQSEEAAYVKADNAASDALEALADQDTALQSEEAAYVSDNNAASDAWEKA 120  
QY 122 ATPIALDVKKTKDKTPVKKERQNVNTLPTTGE 156  
DB 121 ATPIALDVKKTKDKTPVKKERQNVNTLPTTGE 155  
RESULT 6  
AAV71045  
ID AAV71045 standard; Protein; 271 AA.  
XX  
XX AAV71045;  
XX  
XX 29-AUG-2000 (first entry)  
XX  
XX Streptococcus pyogenes strain AP49 partial GRAB protein.  
XX  
XX DE Streptococcus pyogenes strain AP49 partial GRAB protein.  
XX  
XX KW GRAB protein; protein G related alpha2m binding protein; vaccine;  
XX  
XX alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;  
XX  
XX immune response; Streptococcus pyogenes infection.  
XX  
XX OS Streptococcus pyogenes.  
XX  
XX PN WO200026240-A2.  
XX  
XX 11-MAY-2000.  
XX  
XX PD 02-NOV-1999; 99WO-GB03631.  
XX  
XX PF 02-NOV-1999; 99WO-GB03631.  
XX  
XX PR 02-NOV-1998; 98GB-0023975.  
XX

PA (ACTI-) ACTINOVA LTD.  
XX  
XX Bjorck LH, Rasmussen M;  
XX  
XX WPI: 2000-365572/31.  
XX  
XX N-PSDB: AAD00563.  
XX  
XX New alpha2m binding protein for generating a protective immune response  
XX  
XX to group A streptococcus and purifying the binding protein -  
XX  
XX Claim 5; Page 61-62; 67pp; English.  
XX  
XX The patent discloses a new family of proteins termed GRAB (protein G  
XX  
XX related alpha2m binding protein) from Streptococcus pyogenes which have  
XX  
XX the ability to bind alpha2-macroglobulin (alpha2m) and show homology to  
XX  
XX protein G of group G Streptococcus. GRAB protein and peptides derived  
XX  
XX from it are used in vaccine compositions for generating a protective  
XX  
XX immune response against group A Streptococcus. Antibodies against GRAB  
XX  
XX are useful for treating Streptococcus pyogenes infections. The protein  
XX  
XX is also useful for purifying alpha2m from a sample. The present sequence  
XX  
XX is a partial GRAB protein from S. pyogenes strain AP49.  
XX  
XX The protein has alpha2m binding region and is useful in vaccine  
XX  
XX composition.  
XX  
XX Sequence 271 AA:  
SQ  
Query Match 89.3%; Score 694; DB 21; Length 271;  
Best Local Similarity 58.2%; Pred. No. 5.5e-54;  
Matches 156; Conservative 0; Mismatches 0; Indels 112; Gaps 1;  
QY 1 VDSPTEOPRIIPNGGTLTNLGNAPKALRNEERAIDELKKAIEDKATTAIEAAS-- 58  
DB 4 VDSPTEOPRIIPNGGTLTNLGNAPKALRNEERAIDELKKAIEDKATTAIEAASD 63  
QY 59 ----- 58  
DB 64 ALBALDQDALQSEEAAYVSDNNAASDALEALADQDALQSEEAAYVSDNNAAGALRA 123  
QY 59 -----SDALEALADQ 68  
DB 124 LADQDTALQSEEAAYVKADNAASDALEALADQDTALQSEEAAYVKADNAASDALEALADQ 183  
QY 69 TDALQSEEAAYVKADNAASDALEALADQDTALQSEEAAYVSDNNAASDAWEKAATPIALD 128  
DB 184 TDALQSEEAAYVKADNAASDALEALADQDTALQSEEAAYVSDNNAASDAWEKAATPIALD 243  
QY 129 VKTKDKTPVKKERQNVNTLPTTGE 156  
DB 244 VKTKDKTPVKKERQNVNTLPTTGE 271  
RESULT 7  
AAV71043  
ID AAV71043 standard; Protein; 259 AA.  
XX  
XX AAV71043;  
XX  
XX 29-AUG-2000 (first entry)  
XX  
XX Streptococcus pyogenes strain KTL9 partial GRAB protein.  
XX  
XX DE Streptococcus pyogenes strain KTL9 partial GRAB protein.  
XX  
XX KW GRAB protein; protein G related alpha2m binding protein; vaccine;  
XX  
XX alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;  
XX  
XX immune response; Streptococcus pyogenes infection.  
XX  
XX OS Streptococcus pyogenes.  
XX  
XX PN WO200026240-A2.  
XX  
XX 11-MAY-2000.  
XX  
XX PD 02-NOV-1999; 99WO-GB03631.  
XX  
XX PF 02-NOV-1999; 99WO-GB03631.  
XX

PR 02-NOV-1998; 98GB-0023975.  
XX  
XX (ACTI-) ACTINOVA LTD.  
XX  
XX Bjorck LH, Rasmussen M;  
XX  
XX WPI; 2000-365572/31.  
DR N-PSDB; AAD00561.  
XX  
PT New alpha2M binding protein for generating a protective immune response  
PT to group A streptococcus and purifying the binding protein  
XX  
XX  
PS Claim 5; Page 59-60; 67pp; English.  
XX  
XX The patent discloses a new family of proteins termed GRAB (protein G  
CC related alpha2M binding protein) from Streptococcus pyogenes which have  
CC the ability to bind alpha2-macroglobulin (alpha2M) and show homology to  
CC protein G of group G Streptococcus. GRAB protein and peptides derived  
CC from it are used in vaccine compositions for generating a protective  
CC immune response against group A Streptococcus. Antibodies against GRAB  
CC are useful for treating Streptococcus pyogenes infections. The protein  
CC is also useful for purifying alpha2M from a sample. The present sequence  
CC is a partial GRAB protein from S. pyogenes strain KTL9.  
CC The protein has alpha2M binding region and is useful in vaccine  
CC composition.  
XX  
XX Sequence 259 AA:  
SQ  
Query Match 89.1%; Score 692; DB 21; Length 259;  
Best Local Similarity 64.0%; Pred. No. 7.8e-54;  
Matches 153; Conservative 0; Mismatches 2; Indels 84; Gaps 1;  
QY 1 VDSPIEQPRIIPNGTITNLGNAPKALRNERAIDELKKAIEDKEATTAIEAASD 60  
DB 21 VDSPIEQPRIIPNGTITNLGNAPKALRNERAIDELKKAIEDKEATTAIEAASD 80  
QY 61 A----- 61  
DB 81 ALEALADQDALOSEEAAYVQSDNNAASDALEALADQDALOSEEAAYVQSDNNAASD 140  
QY 62 -----LEALADQDALOSEEAAYVQSDNNAASDALEALADQ 96  
DB 141 LADQTDALOSEEAAYVQSDNNAASDALEALADQDALOSEEAAYVQSDNNAASD 200  
QY 97 TDALOSEEAAYVQSDNNAASDALEALADQDALOSEEAAYVQSDNNAASD 155  
DB 201 TDALOSEEAAYVQSDNNAASDALEALADQDALOSEEAAYVQSDNNAASD 259  
RESULT 8  
AA71040  
ID AA71040 standard; peptide; 141 AA.  
XX  
XX AA71040;  
XX  
XX 29-AUG-2000 (first entry)  
XX  
XX Streptococcus pyogenes strain SF370 GRAB protein fragment #4.  
XX  
XX Streptococcus pyogenes strain SF370 GRAB protein fragment #4.  
XX  
XX GRAB protein; protein G related alpha2M binding protein; vaccine;  
KW alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;  
KM immune response; Streptococcus pyogenes infection.  
XX  
XX Streptococcus pyogenes.  
OS  
XX  
XX WO200026240-A2.  
PN  
XX  
XX 11-MAY-2000.  
PD  
XX  
XX 02-NOV-1999; 99WO-GB03631.  
PF  
XX  
XX 02-NOV-1998; 98GB-0023975.  
PR  
XX

PA (ACTI-) ACTINOVA LTD.  
XX  
XX Bjorck LH, Rasmussen M;  
XX  
XX WPI; 2000-365572/31.  
DR  
XX  
XX New alpha2M binding protein for generating a protective immune response  
PT to group A streptococcus and purifying the binding protein  
XX  
XX  
PS Claim 5; Page 57; 67pp; English.  
XX  
XX The patent discloses a new family of proteins termed GRAB (protein G  
CC related alpha2M binding protein) from Streptococcus pyogenes which have  
CC the ability to bind alpha2-macroglobulin (alpha2M) and show homology to  
CC protein G of group G Streptococcus. GRAB protein and peptides derived  
CC from it are used in vaccine compositions for generating a protective  
CC immune response against group A Streptococcus. Antibodies against GRAB  
CC are useful for treating Streptococcus pyogenes infections. The protein  
CC is also useful for purifying alpha2M from a sample. The present sequence  
CC is a fragment of GRAB protein from S. pyogenes strain SF370 corresponding  
CC to residues 34-174. This fragment is devoid of the trans-membrane  
CC and cell wall anchor regions. It is useful in vaccine composition.  
XX  
XX Sequence 141 AA:  
SQ  
Query Match 87.8%; Score 682; DB 21; Length 141;  
Best Local Similarity 100.0%; Pred. No. 2.8e-53;  
Matches 141; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 VDSPIEQPRIIPNGTITNLGNAPKALRNERAIDELKKAIEDKEATTAIEAASD 60  
DB 1 VDSPIEQPRIIPNGTITNLGNAPKALRNERAIDELKKAIEDKEATTAIEAASD 60  
QY 61 ALEALADQDALOSEEAAYVQSDNNAASDALEALADQDALOSEEAAYVQSDNNAASD 120  
DB 61 ALEALADQDALOSEEAAYVQSDNNAASDALEALADQDALOSEEAAYVQSDNNAASD 120  
QY 121 AAPPIALDVKKTKDTPVKK 141  
DB 121 AAPPIALDVKKTKDTPVKK 141  
RESULT 9  
AA71037  
ID AA71037 standard; peptide; 58 AA.  
XX  
XX AA71037;  
XX  
XX 29-AUG-2000 (first entry)  
XX  
XX Streptococcus pyogenes strain SF370 GRAB protein fragment #2.  
XX  
XX Streptococcus pyogenes strain SF370 GRAB protein fragment #2.  
XX  
XX GRAB protein; protein G related alpha2M binding protein; vaccine;  
KW alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;  
KM immune response; Streptococcus pyogenes infection.  
XX  
XX Streptococcus pyogenes.  
OS  
XX  
XX WO200026240-A2.  
PN  
XX  
XX 11-MAY-2000.  
PD  
XX  
XX 02-NOV-1999; 99WO-GB03631.  
PF  
XX  
XX 02-NOV-1998; 98GB-0023975.  
PR  
XX  
XX (ACTI-) ACTINOVA LTD.  
XX  
XX Bjorck LH, Rasmussen M;  
XX  
XX WPI; 2000-365572/31.  
DR  
XX  
XX New alpha2M binding protein for generating a protective immune response

PT to group A streptococcus and purifying the binding protein -  
XX  
PS Claim 2; Page 55; 67pp; English.  
XX  
CC The patent discloses a new family of proteins termed GRAB (protein G  
CC related alpha2m binding protein) from streptococcus pyogenes which have  
CC the ability to bind alpha2-macroglobulin (alpha2m) and show homology to  
CC protein G of group G streptococcus. GRAB protein and peptides derived  
CC from it are used in vaccine compositions for generating a protective  
CC immune response against group A streptococcus. Antibodies against GRAB  
CC are useful for treating streptococcus pyogenes infections. The protein  
CC is also useful for purifying alpha2m from a sample. The present sequence  
CC is a fragment of GRAB protein from S. pyogenes strain SF370 corresponding  
CC to residues 34-91. This fragment is capable of binding alpha2m  
CC and useful in vaccine composition.  
XX  
SQ Sequence 58 AA:  
Query Match 36.7%; Score 285; DB 21; Length 58;  
Best Local Similarity 100.0%; Pred. No. 2e-18;  
Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 VDSPIEQPIIPNGGTLTNLGNAPKALRNBERAIDELKKQAIEDKETTATIAEAS 58  
DB 1 VDSPIEQPIIPNGGTLTNLGNAPKALRNBERAIDELKKQAIEDKETTATIAEAS 58  
RESULT 10  
AAP70493  
ID AAP70493 standard; protein; 448 AA.  
XX  
AC AAP70493;  
XX  
DT 06-MAR-1991 (first entry)  
XX  
DE Protein G.  
XX  
KM Protein G; antibody; Fc receptor;  
XX  
OS Streptococcus lancefield Group G strain.  
XX  
FH Key Location/Qualifiers  
FT Active-site 228..297  
FT Active-site /label-active site B1  
FT Active-site 298..352  
FT /label-active site B2  
XX  
PN WO8705025-A.  
XX  
PD 27-AUG-1987.  
XX  
PF 17-FEB-1987; 87WO-US00329.  
XX  
PR 22-APR-1986; 86US-0854997.  
PR 14-FEB-1986; 86US-0829354.  
XX  
PA (GENE-) GENEX CORP.  
PA (FAHN-) FAHNESTOCK S R.  
XX  
PI Fahnestock S;  
XX  
DR WPI; 1987-250197/35.  
DR N-PSDB; AAN70811.  
XX  
PT Cloned Protein G gene - used for producing Protein G for  
PT detection and purification. of antibodies and treatment of diseases  
XX  
PS Disclosure; Fig. 3; 68pp; English.  
XX  
CC Protein G expressed by inserting the gene into an expression  
CC vector. A second vector may also be used as a cryptic helper plasmid  
CC to stably maintain the first plasmid in the host cell. Bacterial Fc  
CC receptors such as protein G can be used to detect and purify

CC antibodies, and in the treatment of disease. Fc receptors are useful  
CC to purify antibodies to be used in the purification of protein drugs and  
CC as therapeutics. High levels of Protein G can be obd. in conditions  
CC favourable for isolation, using a non-pathogenic host. Suitable  
CC cloning vectors are lambda gt11, M13mp9 and pCX1066.  
XX  
SQ Sequence 448 AA:  
Query Match 24.2%; Score 188; DB 8; Length 448;  
Best Local Similarity 38.9%; Pred. No. 1e-08;  
Matches 61; Conservative 20; Mismatches 60; Indels 16; Gaps 6;  
OY 1 VDSPIEQPIIPNGGTLTNLGNAPKALRNBERAIDELKKQAIEDKETTATIAEASD 60  
DB 34 VDSPIEDPTIIRNGGELTNLGNSETTLALRNEESATADLTAAAVADTVAAAAAEVACAA 93  
OY 61 ALEALADDTDAQSEEAAYVKADN--AASDALEAL--ADQTDALQSEAEVVGSDNMA- 114  
DB 94 AWEAALA-AADALAKAKADLKEFNKYGVSDYKLNILNNKTYEGIKDLOAOVESAKKAR 152  
OY 115 -SDANE-----KAATPIALDVKKTK--DTKPEVVKKE 142  
DB 153 ISEATDGLSDFLKSGTPAEDTVKKSIEIAEAKYLAANRE 189  
RESULT 11  
AAP95030  
ID AAP95030 standard; protein; 448 AA.  
XX  
AC AAP95030;  
XX  
DT 04-JUL-1990 (first entry)  
XX  
DE Protein G.  
XX  
KM Protein G; immunoglobulin; Fc receptor; ds.  
XX  
OS Streptococcus sp.  
XX  
PN WO8810306-A.  
XX  
PD 29-DEC-1988.  
XX  
PF 20-JUN-1988; 88WO-US02084.  
XX  
PR 19-JUN-1987; 87US-0063959.  
XX  
PA (GENE-) GENEX CORP.  
XX  
PI Fahnestock SR;  
XX  
DR WPI; 1989-023848/03.  
DR N-PSDB; AAN91093.  
XX  
PT Cloned protein G variant genes -  
PT expressing proteins having immunoglobulin-binding properties of  
PT protein G and derived from streptococcus sp.  
XX  
PS Disclosure; 116pp; English.  
XX  
CC Protein G of non-pathogenic streptococcus and variants may be isolated.  
CC useful as bacterial Fc receptors eg in purification and detection of Abs.  
CC screening of hybridoma clones and treatment of disease.  
XX  
SQ Sequence 448 AA:  
Query Match 24.2%; Score 188; DB 10; Length 448;  
Best Local Similarity 38.9%; Pred. No. 1e-08;  
Matches 61; Conservative 20; Mismatches 60; Indels 16; Gaps 6;  
OY 1 VDSPIEQPIIPNGGTLTNLGNAPKALRNBERAIDELKKQAIEDKETTATIAEASD 60  
DB 34 VDSPIEDPTIIRNGGELTNLGNSETTLALRNEESATADLTAAAVADTVAAAAAEVACAA 93

OY	61	ALEALADDTDALQSEAEAAVVKADN--AASDALEAL--ADOTDALQSEAEAVVOSDMA- 114
Dd	94	AMEAAA-AADLAKKKAIDALEFKNKYGVSDYYKKLLINNAKTVEGIKDLQAQVESAKRAR 152
OY	115	-SDAME-----KATPIALDVKKTK--DTRPVYKKE 142
Dd	153	ISEATDGLSDFLKSTQPAEDTYKSTIELAEKVLANRE 189
 RESULT 12		
ID	AAR07013	standard; protein: 448 AA.
XX	AC	AAR07013;
XX	AD	
XX	DE	17-JAN-1991 (first entry)
XX	DM	Protein G variant with two active sites.
XX	KM	Immunoglobulin.
OS	SM	Streptococcus sp. Lancefield Group G strain.
XX	FH	Key
FT	FT	Active-site            Location/Qualifiers
FT	FT	/label=B1
FT	FT	Active-site            228..282
FT	FT	/label=B2
FT	FT	Active-site            298..352
FT	FT	/label=B2
XX	PN	US4956296-A.
PD	PD	11-SEP-1990.
XX	PF	20-JUN-1988;      88US-0209236.
XX	PR	20-JUN-1988;      88US-0209236.
PR	PR	14-FEB-1986;      86US-08329354.
PR	PR	23-APR-1986;      86US-0854887.
PR	PR	17-FEB-1987;      87MO-US00328.
PR	PR	19-JUN-1987;      87US-0063939.
XX	PA	(GENE-) GENEX CORP.
XX	PI	Fahnestock SR.
XX	DR	WPI: 1990-297491/39.
XX	DR	N-PSDB: AAO06018.
PT	PT	Recombinant Protein G variants - obid. using a cloned gene
PT	PT	encoding Protein G from Streptococcus sp., used for binding
PT	PT	immunoglobulin.
XX	PS	Disclosure: Fig 8a-c: 48pp; English.
CC	CC	Fragments and variants of the sequence are claimed esp. where
CC	CC	incorporated into a non-pathogenic host eg. E.coli, and expressed
CC	CC	at high levels.
CC	CC	The variants have a higher binding efficiency and capacity for
CC	CC	immunoglobulin, and may be used for purifying, detecting and
CC	CC	isolating antibodies.
XX	SQ	Sequence      448 AA;
 Query Match            24.2%; Score 188; DB 11; Length 448;		
Best Local Similarity   38.9%; Pred. No. 1e-08;		
Matches    61; Conservative   20; Mismatches   60; Indels   16; Gaps    6;		
OY	1	VDSPIEOPRIIPNGTILNLGNAPDKALNRNEERAIDELTKQATEDEKATAIEMASD 60
Dd	34	VDSPIEDPIITINGSELITNLGNSETTIALNRNESATADLTAAAYADVIAAAAAENAGCA 93
OY	61	ALEALADDTDALQSEAEAAVVKADN--AASDALEAL--ADOTDALQSEAEAVVOSDMA- 114

Db	94	AMEAAA-AADALAAKKAALAKKAFNNKGYSDYYKKNLINNAKVEGIGIKDLQAQVYESAKKAR	152			
QY	115	-SDAME-----KAATPIALDVYKTK--DTKPVVKE	142			
Db	153	ISEATFDGLSDFLKSOTPAEDFVKSIELAAKVLNRE	189			
RESULT 13						
ID	AAAI0004	standard; Protein: 448 AA.				
AC	AAAI0004;					
XX						
XX	13-MAR-1991	(first entry)				
XX						
DE	Streptococcus GX7809	protein G.				
XX						
KW	Immunoglobulins; Ig.					
XX						
OS	Streptococcus sp GX7809.					
XX						
FT	Key	Location/Qualifiers				
FT	Active-site	228..282				
FT	Active-site	/label- Active site B1				
FT		298..352				
FT		/label- Active site B2				
XX						
PN	USA977247-A.					
XX						
PD	11-DEC-1990.					
XX						
PF	19-MAY-1989;	89US-0354264.				
XX						
PR	19-MAY-1989;	89US-0354264.				
PR	14-FEB-1986;	86US-0829354.				
PR	23-APR-1986;	86US-0834887.				
PR	17-FEB-1987;	87MO-US00329.				
PR	19-JUN-1987;	87US-0063959.				
PR	20-JUN-1988;	88US-0209236.				
XX						
PA	(GENE-) GENEX CORP.					
XX						
PI	Fahnestoc SR, Lee T, Wroble MH;					
XX						
DR	WPI: 1991-006758/01.					
DR	Q-PSDB: Q10001.					
XX						
PT	Immoblised protein G variants - used for detection, isolation					
PT	and purification. immunoglobulin(s) and immunoglobulin fragments					
XX						
PS	Disclosure: Fig 8; 52pp: English.					
XX						
CC	Protein G gene product may be modified allowing the variant to be					
CC	immobilised and exhibit different binding profiles. The bound					
CC	protein is useful in purification and detection of Igs and fragments.					
XX						
XX						
SO	Sequence	448 AA:				
QY	Query Match	24.2%; Score 188; DB 12; Length 448;				
Db	Best Local Similarity	38.9%; Pred. NO. 1e-08;				
	Matches	61; Conservative	20; Mismatches	60; Indels	16; Gaps	6.
QY	1	VDSPIEQPIITNGTTLGNAPKIALRNREERAIDELKQATIEDKREATTALAEASSD	60			
Db	34	VDSPIEQPIITNGGTLNGLNSETTLALRNDESATALPAAVAADTVAAAAAENAGAA	93			
QY	61	ALEAALADOTDALQSEEAAYVAKADN--AASDALEAL--ADQTDALQSEEAAYVQSDNAA-	114			
Db	94	AMEAAA-AADALAAKKAALAKKAFNNKGYSDYYKKNLINNAKVEGIGIKDLQAQVYESAKKAR	152			
QY	115	-SDAME-----KAATPIALDVYKTK--DTKPVVKE	142			

```

Db      153 ISEATDGLSDFLKSGTPAEDTVKSIELAEKVLNRE 189
RESULT 14
AAR53290
ID      AAR53290 standard; Protein; 448 AA.
XX
XX      AAR53290;
AC
XX
XX      06-JAN-1995 (first entry)
DT
XX
XX      Streptococcus Protein G derived from strain GX7809.
DE
XX
XX      Streptococcus Protein G; variant; IgG binding activity;
KW      immunoglobulin; Lancefield Group G; bacterial Fc receptor.
XX
OS      Streptococcus sp. GX7809.
XX
XX      Key      Location/Qualifiers
FH      Active-site 228..282
FT      /label=B1
FT      Region      283..297
FT      /label=B
FT      /note="Linking region"
FT      Active-site 298..352
FT      /label=B2
XX
XX      US5312901-A.
PN
XX
XX      17-MAY-1994.
PD
XX
XX      14-FEB-1986; 86US-0829354.
PF
XX
XX      14-FEB-1986; 86US-0829354.
PR
XX      14-FEB-1986; 86US-0829354.
PR      23-APR-1986; 86US-0854887.
PR      19-JUN-1987; 87US-0063959.
PR      20-JUN-1988; 86US-0209236.
PR      19-JUN-1990; 90US-0540169.
PR      21-APR-1992; 92US-0871539.
XX
XX      (PHMA ) PHARMACIA LKB BIOTECHNOLOGY AB.
PA
XX
XX      Fahnstock SR;
PI
XX
XX      WPI; 1994-159179/19.
DR
XX      N-PSDB; AA064644.
DR
XX
XX      New recombinant streptococcal protein G variants - useful for
PT      antibody detection and purification and for therapy
PT
XX
XX      Example 2; Fig 3 and Fig 8; 48pp; English.
PS
XX
XX      A 1.9kb HindIII fragment containing the entire coding sequence for
CC      Protein G was isolated from Streptococcus GX7809. The protein G has
CC      IgG-binding activity which has been localised to the B repeating
CC      structure. Streptococcal Protein G variants comprising the B domains
CC      are claimed.
XX
XX
SQ      Sequence 448 AA:

Query Match      24.2%; Score 188; DB 15; Length 448;
Best Local Similarity 38.9%; Pred. No. 1e-08;
Matches 61; Conservative 20; Mismatches 60; Indels 16; Gaps 6;

```

```

Db      153 ISEATDGLSDFLKSGTPAEDTVKSIELAEKVLNRE 189
RESULT 15
AAR07014
ID      AAR07014 standard; protein; 593 AA.
XX
XX      AAR07014;
AC
XX
XX      17-JAN-1991 (first entry)
DT
XX
XX      Protein G variant with three active sites.
DE
XX
XX      immunoglobulin.
KW
XX
OS      Streptococcus sp. Lancefield Group G strain.
XX
XX      Key      Location/Qualifiers
FH      Active-site 303..372
FT      /label=B1
FT      Active-site 373..427
FT      /label=B3
FT      Active-site 443..497
FT      /label=B2
XX
XX      US4956296-A.
PN
XX
XX      11-SEP-1990.
PD
XX
XX      20-JUN-1988; 88US-0209236.
PF
XX
XX      20-JUN-1988; 88US-0209236.
PR
XX      20-JUN-1988; 88US-0209236.
PR      14-FEB-1986; 86US-0829354.
PR      23-APR-1986; 86US-0854887.
PR      17-FEB-1987; 87MO-US00329.
PR      19-JUN-1987; 87US-0063959.
XX
XX      (GENE-) GENEX CORP.
PA
XX
XX      Fahnstock SR;
PI
XX
XX      WPI; 1990-297491/39.
DR
XX      N-PSDB; AA06019.
DR
XX
XX      Recombinant Protein G variants - obt'd. using a cloned gene
PT      encoding Protein G from Streptococcus sp., used for binding
PT      immunoglobulin.
XX
XX      Disclosure; Fig 9; 48pp; English.
PS
XX
XX      Fragments and variants of the sequence are claimed esp. where
CC      incorporated into a non-pathogenic host eg. E.coli, and expressed
CC      at high levels.
CC      The variants have a higher binding efficiency and capacity for
CC      immunoglobulin, and may be used for purifying, detecting and
CC      isolating antibodies.
XX
XX
SQ      Sequence 593 AA:

Query Match      24.1%; Score 187; DB 11; Length 593;
Best Local Similarity 38.9%; Pred. No. 1.8e-08;
Matches 61; Conservative 20; Mismatches 60; Indels 16; Gaps 6;

```



```

Db      153  ISEATDGLSDPLKSGTQPAEDTVKSIELAEKVLNRE 189
RESULT 16
AAR62944
ID      AAR62944 standard; Protein; 593 AA.
AC      AAR62944;
XX
XX      10-JAN-1995 (first entry)
XX
XX      Streptococcus Protein G derived from strain GX7805.
XX
XX      Streptococcus Protein G; variant; IgG binding activity;
XX      immunoglobulin; Lancefield Group G; bacterial Fc receptor.
XX
XX      Streptococcus sp. GX7805.
XX
XX      Key      Location/Qualifiers
XX      FH      Domain      106..140
XX      FT      /label= A1
XX      FT      141..178
XX      FT      /label= a1
XX      FT      /note= "linking region"
XX      FT      179..215
XX      FT      /label= A2
XX      FT      216..253
XX      FT      /label= a2
XX      FT      /note= "linking region"
XX      FT      254..290
XX      FT      /label= A3
XX      FT      269
XX      FT      /note= "corresponds to GGA codon"
XX      FT      303..357
XX      FT      /label= B1
XX      FT      /note= "Involved in IgG binding activity"
XX      FT      358..372
XX      FT      /label= b
XX      FT      /note= "linking region"
XX      FT      373..427
XX      FT      /label= B3
XX      FT      /note= "Involved in IgG binding activity"
XX      FT      376
XX      FT      /note= "corresponds to CCF codon"
XX      FT      428..442
XX      FT      /label= b
XX      FT      /note= "linking region"
XX      FT      443..497
XX      FT      /label= B2
XX      FT      /note= "Involved in IgG binding activity"
XX      FT      466
XX      FT      /note= "corresponds to ACT codon"
XX      FT      531..535
XX      FT      /label= C1
XX      FT      536..540
XX      FT      /label= C2
XX      FT      541..545
XX      FT      /label= C3
XX      FT      546..550
XX      FT      /label= C4
XX      FT      551..555
XX      FT      /label= C5
XX      FT      Misc-difference 592
XX      FT      /note= "corresponds to GAA codon"
XX
XX      US5312901-A.
XX      PN
XX      17-MAY-1994.
XX      PD
XX      14-FEB-1986; 86US-0829354.
XX      PF
XX      14-FEB-1986; 86US-0829354.
XX      PR

```

```

PR      23-APR-1986; 86US-0854887.
PR      19-JUN-1987; 87US-0063959.
PR      20-JUN-1988; 88US-0209226.
PR      19-JUN-1990; 90US-0540169.
PR      21-APR-1992; 92US-0871539.
XX
XX      (PHAA ) PHARMACIA LKB BIOTECHNOLOGY AB.
XX
XX      Fahnestock SR;
XX
XX      WPI; 1994-159179/19.
XX      DR      N-PSDB; AA075036.
XX
XX      New recombinant streptococcal protein G variants - useful for
XX      antibody detection and purification and for therapy
XX
XX      Example 5; Fig 9; 48pp; English.
XX
XX      A 2.4kb HindIII fragment containing the entire coding sequence for
XX      Protein G was isolated from Streptococcus GX7805 using the 1.9kb
XX      Protein G coding sequence from Streptococcus GX7809. The Protein G
XX      has IgG-binding activity which has been localised to the B repeating
XX      structure. Streptococcal Protein G variants comprising the B domains
XX      are claimed.
XX
XX      Sequence 593 AA;
XX
XX      Query Match      24.1%; Score 187; DB 15; Length 593;
XX      Best local Similarity 38.9%; Pred. No. 1.8e-08;
XX      Matches 61; Conservative 20; Mismatches 60; Indels 16; Gaps 6;
XX
XX      1 VDSPTEQPRITIPNGGTNLNLGNAPKALNBERAIDELKKAIIEDEATTATEAASD 60
XX      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX      34 VDSPEDPTIRNGEELNLGNSETTLALRNESATADLTPAAVADTVAAAAAENAGAA 93
XX
XX      61 ALAALADDTDALQSEEAAYVRADN--AASDALEAL--ADDTDALQSEEAAYVOSDNAA- 114
XX      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX      94 AWEAAA-AAADALAKAKADALKEPNKYGVSDYYKNLNNAKTVEGVKDDQAVDSAKKAR 152
XX
XX      115 -SDAME-----KAATPIALDYKTK--DTKPVYKKE 142
XX      |:|:| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:|
XX      153 ISEATDGLSDPLKSGTQPAEDTVKSIELAEKVLNRE 189
XX
RESULT 17
AAR10005
ID      AAR10005 standard; protein; 594 AA.
XX
XX      AAR10005;
XX
XX      13-MAR-1991 (first entry)
XX
XX      Streptococcus GX7805 protein G.
XX
XX      Immunoglobulins; Ig.
XX
XX      Streptococcus sp GX7805.
XX
XX      Key      Location/Qualifiers
XX      FH      Active-site 304..358
XX      FT      /label= Active site B1
XX      FT      374..428
XX      FT      /label= Active site B3
XX      FT      444..498
XX      FT      /label= Active site B2
XX
XX      US4977247-A.
XX      PN
XX      11-DEC-1990.
XX      PD
XX      19-MAY-1989; 89US-0354264.
XX      PF
XX      19-MAY-1989; 89US-0354264.
XX      PR

```

PR 14-FEB-1986; 86US-0829354.  
 PR 23-APR-1986; 86US-0854887.  
 PR 17-FEB-1987; 87WO-US00329.  
 PR 19-JUN-1987; 87US-0063959.  
 PR 20-JUN-1988; 88US-0209236.  
 XX  
 PA (GENE-) GENEX CORP.  
 XX  
 PI Fahnestock SR, Lee T, Wroble MH;  
 DR WPI; 1991-006758/01.  
 DR Q-PSDB; Q10002.  
 XX  
 PT Immobilised protein G variants - used for detection, isolation  
 PT and purification. Immunoglobulin(s) and immunoglobulin fragments  
 PS Disclosure; Fig 9; 52pp; English.  
 CC  
 CC Protein G gene product may be modified allowing the variant to  
 CC be immobilised and exhibit different binding profiles. The bound  
 CC protein is useful in purification and detection of Igs and fragments.  
 XX  
 SQ Sequence 594 AA;

Query Match 24.1%; Score 187; DB 12; Length 594;  
 Best Local Similarity 38.9%; Pred. No. 1.8e-08;  
 Matches 61; Conservative 20; Mismatches 60; Indels 16; Gaps 6;  
 OY 1 VDSPIDPPIPIRNGGTITNLGNAPKRLRNBERAIDELKKAQIEDKKEATTAEASD 60  
 DB 34 VDSPIDPPIPIRNGGTITNLGNSETTLALRNESATADITAAVADTVAAAAAENAGAA 93  
 OY 61 ALEALADQTDALQSEEAAYKADN--AASDALEAL--ADQTDALQSEEAAYQSDNAA- 114  
 DB 94 AWEAAR--AADALAKAKADALKEFNKYGVSDYKKNLNNAKTVEGVVDLQAVVESAKKAR 152  
 OY 115 -SDPAWE-----KAATPIALDVKKTK--DTKKYVKKKE 142  
 DB 153 ISEATDGLSDFLKSQTPAEDTVASIELAEKVLANRE 189

RESULT 18  
 AAY57611  
 ID AAY57611 standard; protein; 103 AA.

AC AAY57611;  
 DT 10-MAR-2000 (first entry)  
 DE Streptococcus strain G 148 protein.

XX Streptococcus strain G 148; protein G'; protein G primer; liposome;  
 KW liposomal delivery complex; connecting moiety; antibody; Fc region;  
 KW diagnosis; therapeutic agent; antibiotic; antidiagnostic; antiviral;  
 KW antitumorigenic; cytokine; hormone; imaging agent; neurotransmitter;  
 KW stimulant; cytotoxic agent; malignant cell.

XX Streptococcus sp.  
 OS  
 XX WO9959643-A2.  
 PN  
 XX 25-NOV-1999.  
 PD  
 XX 19-MAY-1999; 99WO-US11177.  
 PF  
 XX 20-MAY-1998; 98US-0086347.  
 PR  
 XX (SDGS-) SDG INC.

PA Lau JR;  
 PI  
 XX WPI; 2000-062383/05.  
 DR N-PSDB; AA247931.  
 XX

XX New liposome constructs comprising a liposome connected to an antibody,  
 PT used, e.g. for delivery of cytotoxic agents to malignant cells  
 PT  
 XX  
 PS Disclosure; Fig 1; 32pp; English.

CC The present invention describes a construct for connecting an antibody  
 CC or antibody fragment to a liposome. The construct comprises protein G'  
 CC and a linking moiety for connecting the protein G' to the liposome.  
 CC Also described are: (1) a liposomal delivery vehicle comprising: (a) a  
 CC liposome; and (b) a connecting moiety connected to the liposome, which  
 CC specifically binds the Fc region of an antibody, for connecting the  
 CC antibody to the liposome; and (2) forming a liposomal carrier comprising:  
 CC (a) providing liposomal components having binding moieties, for forming  
 CC a liposome; (b) providing a construct comprising a linking moiety and a  
 CC connecting moiety bound together; and (c) combining the liposomal  
 CC components and the construct and sonicating the combination so that the  
 CC binding moieties are exposed on a surface of the liposome, for binding to  
 CC the linking moiety, and to facilitate the binding between them. The  
 CC products can be used for the delivery of diagnostic or therapeutic  
 CC agents. The liposomes may contain or may be associated with a diagnostic  
 CC or therapeutic agent, e.g. antibiotics, antidepressants,  
 CC antitumorigenic, antivirals, cytokines, hormones, imaging agents,  
 CC neurotransmitters, or stimulants. They can be used particularly for the  
 CC delivery of cytotoxic agents to malignant cells. The protein G'  
 CC connecting moiety provides a liposomal delivery complex having improved  
 CC targeting efficiency. As a result of the binding between protein G' and  
 CC the Fc region of antibodies, protein G' shields the Fc regions of the  
 CC attached antibodies from non-specific binding to cell surfaces, other  
 CC proteins, and anatomical structures. The present sequence represents  
 CC an unidentified protein encoded by the same sequence which encodes  
 CC protein G'.

SQ Sequence 103 AA;  
 Query Match 18.3%; Score 142; DB 21; Length 103;  
 Best Local Similarity 57.9%; Pred. No. 2.1e-05;  
 Matches 33; Conservative 3; Mismatches 21; Indels 0; Gaps 0;

OY 1 VDSPIDPPIPIRNGGTITNLGNAPKRLRNBERAIDELKKAQIEDKKEATTAEAA 57  
 DB 34 VDSPIDPPIPIRNGGTITNLGNSETTLALRNESATADITAAVADTVAAAAAENAA 90

RESULT 19  
 AAP94785  
 ID AAP94785 standard; protein; 235 AA.

AC AAP94785;  
 DT 04-JUL-1990 (first entry)  
 DE Protein G variant.

XX Protein G; immunoglobulin; Fc receptor; ds.

XX Streptococcus sp.  
 OS  
 XX WO8810306-A.  
 PN  
 XX 29-DEC-1988.  
 PD  
 XX 20-JUN-1988; 88WO-US02084.  
 PF  
 XX 19-JUN-1987; 87US-0063959.  
 PR  
 XX (GENE-) GENEX CORP.

PA Fahnestock SR;  
 PI  
 XX WPI; 1989-023848/03.  
 DR N-PSDB; AAN91099.  
 XX

PT Cloned protein G variant genes -  
 PT expressing proteins having immunoglobulin-binding properties of  
 PT protein G and derived from Streptococcus sp.  
 XX  
 PS Claim 17; Page 88; 116pp; English.  
 XX  
 CC Gene for protein G variant of non-pathogenic streptococcus sp. allowing  
 CC isolation of the protein and variants, useful as bacterial Fc receptors  
 CC eg in purification and detection of Abs., screening of hybridoma clones  
 CC and treatment of disease.  
 CC  
 SQ Sequence 235 AA;  
 Query Match 17.9%; Score 139; DB 10; Length 235;  
 Best Local Similarity 25.2%; Pred. No. 0.00011;  
 Matches 58; Conservative 24; Mismatches 64; Indels 84; Gaps 8;  
 OY 4 PIEDPTIRNGELTNLGNAPKALRNBERAIDE-----LKKQAIEDKEAT 51  
 DB 4 PIEDPTIRNGELTNLGNSETTLALRNESATAGYPLPKTDYKYLINGKLGKETT 63  
 OY 52 TAIEAASSD-----ALEALADQTDALQ---SEEAAYKAD-----N 84  
 DB 64 EAVDAATAEKVFQRYANDNGVDGMTYDDATKTFVTEKPEVIDASELPVAVTTYKLVIN 123  
 OY 85 AASDALEALADQTDALQSEEA-EVYOSDNAAADAW----- 118  
 DB 124 GKTLKGETTTKAVDAETAEKAFQRYANDNGVDGVTYDDATKTFVTEMTVEVPRGDAPT 183  
 OY 119 --EK-----AATPIALDVKKTKDT-----KPVYKKEERQNVNT 149  
 DB 184 EPEKPEASIPVLVLPATPIAKDKKDDTKKDAKKPEAKKDDAKKAET 233  
 RESULT 20  
 AAR07004  
 ID AAR07004 standard; protein; 235 AA.  
 XX  
 AC AAR07004;  
 XX  
 DT 17-JAN-1991 (first entry)  
 XX  
 DE Protein G variant.  
 XX  
 KW Immunoglobulin.  
 XX  
 OS Streptococcus sp. Lancefield Group G strain.  
 XX  
 PN US4956296-A.  
 XX  
 PD 11-SEP-1990.  
 XX  
 PF 20-JUN-1988; 88US-0209236.  
 XX  
 PR 20-JUN-1988; 88US-0209236.  
 PR 14-FEB-1986; 86US-0829354.  
 PR 23-APR-1986; 86US-0854887.  
 PR 17-FEB-1987; 87WO-US00329.  
 PR 19-JUN-1987; 87US-0063959.  
 XX  
 PA (GENE-) GENEX CORP.  
 XX  
 PI Fahnestock SR;  
 XX  
 DR WPI; 1990-297491/39.  
 DR N-PSDB; AAC06009.  
 XX  
 PT Recombinant protein G variants - obtd. using a cloned gene  
 PT encoding protein G from Streptococcus sp., used for binding  
 PT immunoglobulin.  
 XX  
 PS Disclosure; Column 9-16; 48pp; English.  
 XX

CC Sequence may be incorporated into a non-pathogenic host eg. E.coli,  
 CC where they may be expressed at high levels. The proteins have a  
 CC higher binding efficiency and capacity for immunoglobulin, and may  
 CC be used for purifying, detecting and isolating antibodies.  
 CC  
 SQ Sequence 235 AA;  
 Query Match 17.9%; Score 139; DB 11; Length 235;  
 Best Local Similarity 25.2%; Pred. No. 0.00011;  
 Matches 58; Conservative 24; Mismatches 64; Indels 84; Gaps 8;  
 OY 4 PIEDPTIRNGELTNLGNAPKALRNBERAIDE-----LKKQAIEDKEAT 51  
 DB 4 PIEDPTIRNGELTNLGNSETTLALRNESATAGYPLPKTDYKYLINGKLGKETT 63  
 OY 52 TAIEAASSD-----ALEALADQTDALQ---SEEAAYKAD-----N 84  
 DB 64 EAVDAATAEKVFQRYANDNGVDGMTYDDATKTFVTEKPEVIDASELPVAVTTYKLVIN 123  
 OY 85 AASDALEALADQTDALQSEEA-EVYOSDNAAADAW----- 118  
 DB 124 GKTLKGETTTKAVDAETAEKAFQRYANDNGVDGVTYDDATKTFVTEMTVEVPRGDAPT 183  
 OY 119 --EK-----AATPIALDVKKTKDT-----KPVYKKEERQNVNT 149  
 DB 184 EPEKPEASIPVLVLPATPIAKDKKDDTKKDAKKPEAKKDDAKKAET 233  
 RESULT 21  
 AAR10011  
 ID AAR10011 standard; Protein; 265 AA.  
 XX  
 AC AAR10011;  
 XX  
 DT 13-MAR-1991 (first entry)  
 XX  
 DE Type 4 GX7809 protein G variant.  
 XX  
 KW Immunoglobulins; Ig.  
 XX  
 OS Streptococcus sp GX7809.  
 XX  
 PN US4977247-A.  
 XX  
 PD 11-DEC-1990.  
 XX  
 PF 19-MAY-1989; 89US-0354264.  
 XX  
 PR 19-MAY-1989; 89US-0354264.  
 PR 14-FEB-1986; 86US-0829354.  
 PR 23-APR-1986; 86US-0854887.  
 PR 17-FEB-1987; 87WO-US00329.  
 PR 19-JUN-1987; 87US-0063959.  
 PR 20-JUN-1988; 88US-0209236.  
 XX  
 PA (GENE-) GENEX CORP.  
 XX  
 PI Fahnestoc SR, Lee T, Wroble MH;  
 XX  
 DR WPI; 1991-006758/01.  
 DR Q-PSDB; Q10007.  
 XX  
 PT Immobilised protein G variants - used for detection, isolation  
 PT and purificn. Immunoglobulin(s) and immunoglobulin fragments  
 XX  
 PS Disclosure; Column 11; 52pp; English.  
 XX  
 CC Protein G variant product carries active regions B1 and B2. It may  
 CC be immobilised and exhibits different binding profiles. The bound  
 CC protein is useful in purification and detection of Igs and fragments.  
 XX  
 SQ Sequence 265 AA;

[illegible]

RESULT 22  
AAR53294  
ID AAR53294 standard; Protein: 265 AA.

XX	AAK53294;
XX	06-JAN-1995 (first entry)
XX	
DE	IgG-binding Streptococcus Protein G variant.
XX	
KW	Streptococcus Protein G; variant; IgG binding activity;
XX	Immunoglobulin; Lancefield Group G; bacterial Fc receptor.
OS	Streptococcus sp.
XX	
FH	Key
FT	Location/Qualifiers
FT	1..30
FT	/label= secretion_sequence
FT	31..265
FT	/label= protein G variant
FT	/note= "claimed without the secretion sequence"
XX	
PN	US5312901-A.
XX	
PD	17-MAY-1994.
XX	
PE	14-FEB-1986; 86US-0829354.
XX	
PR	14-FEB-1986; 86US-0829354.
PR	23-APR-1986; 86US-0854887.
PR	19-JUN-1987; 87US-0063959.
PR	20-JUN-1988; 88US-0209236.
PR	19-JUN-1990; 90US-0540169.
PR	21-APR-1992; 92US-0871539.
XX	
PA	(PHAA ) PHARMACIA LKB BIOTECHNOLOGY AB.
XX	
PI	Fahnestock SR;
XX	
DR	WPI, 1994-159179/19.
XX	
DR	N-PSDB; AAO64648.
XX	
PT	New recombinant streptococcal protein G variants - useful for
PT	antibody detection and purification and for therapy
XX	
PS	Claim 3; Column 46; 48pp; English.
XX	
CC	Protein G isolated from Streptococcus has IgG-binding activity which
CC	has been localised to the B repeating structure (see AAK53290).
CC	The sequence AAK53294 represents a claimed Streptococcal Protein G
CC	variant comprising the B domain active site and retaining
CC	IgG-binding activity.

[illegible]

RESULT 23	
AAV71038	
ID	AAV71038 standard; peptide; 28 AA

Query Match 16.6%; Score 129; DB 21; Length 28;  
Best Local Similarity 100.0%; Pred. No. 6e-05;  
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 59 SDALEALADQTDALQSEAAVVKADNMA 86  
DB 1 SDALEALADQTDALQSEAAVVKADNMA 28

RESULT 24  
AA71036  
ID AA71036 standard; peptide; 23 AA.  
XX  
AC AA71036;  
XX  
DT 29-AUG-2000 (first entry)  
XX  
DE Streptococcus pyogenes strain SF370 GRAB protein fragment #1.  
XX  
KM GRAB protein; protein G related alpha2M binding protein; vaccine;  
KM alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;  
KM Immune response; Streptococcus pyogenes infection.  
XX  
OS Streptococcus pyogenes.  
XX  
PN WO200026240-A2.  
XX  
PD 11-MAY-2000.  
XX  
PF 02-NOV-1999; 99MO-GB03631.  
XX  
PR 02-NOV-1998; 98GB-0023975.  
XX  
PA (ACT-) ACTINOVA LTD.  
XX  
PI Bjorck LH, Rasmussen M;  
XX  
DR WPI; 2000-365572/31.  
XX  
PT New alpha2M binding protein for generating a protective immune response  
to group A streptococcus and purifying the binding protein -  
XX  
PS Claim 1; Page 55; 67pp; English.  
XX  
CC The patent discloses a new family of proteins termed GRAB (protein G  
related alpha2M binding protein) from Streptococcus pyogenes which have  
the ability to bind alpha2-macroglobulin (alpha2M) and show homology to  
protein G of group G Streptococcus. GRAB protein and peptides derived  
from it are used in vaccine compositions for generating a protective  
immune response against group A Streptococcus. Antibodies against GRAB  
are useful for treating Streptococcus pyogenes infections. The protein  
is also useful for purifying alpha2M from a sample. The present sequence  
is a fragment of GRAB protein from S. pyogenes strain SF370 corresponding  
to residues 34-56. This fragment is capable of binding alpha2M  
and useful in vaccine composition.  
XX  
CC  
XX  
SQ Sequence 23 AA:

Query Match 15.4%; Score 120; DB 21; Length 23;  
Best Local Similarity 100.0%; Pred. No. 0.0003;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VDSPRIOPRIIPNGTITNLGN 23  
DB 1 VDSPRIOPRIIPNGTITNLGN 23

RESULT 25  
AAG38475  
ID AAG38475 standard; Protein; 166 AA.  
XX  
AC AAG38475;  
XX

DT 18-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 47472.  
XX  
KM Protein identification: signal transduction pathway; metabolic pathway;  
KM hybridisation assay; genetic mapping; gene expression control; promoter;  
KM termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 23-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.

```
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144326.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 04-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.

PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0158293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 18-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 14.9%; Score 116; DB 21; Length 166;
Best Local Similarity 26.3%; Pred. No. 0.0078;
Matches 41; Conservative 25; Mismatches 62; Indels 28; Gaps 5;

QY 12 PNGGTLTNLIGNP-----EKLALRNEERAIDELKQALIEDKEATTALIEASSDALEALA 66
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 8 PIESDMQNEGSPVPMKPVSENVYAKENNTESGEKQQTVAETTFETTSVAKKEFPVEPTK 67
      : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 67 DQTDALQSE-----EAAVY---KADNAASDALEALADQTDALQSEEAAYVOS 110
      : | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 68 ETPPAVQPEVAAVESSSADAGEAAVVAPEKEVEMATENAEMAKVEAVVAABEKEVEV--- 124
      : | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 111 DNAASDAEMKA-ATPIALDVKKTKDKPKPVKKKEEQ 145
      | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 125 ---AVEAEKKAEPYKAEPEYKAEPEPKESKO 157
      : | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 26
AAV71048
ID AAV71048 standard: peptide; 25 AA.
XX
AC AAV71048;
```

XX 29-AUG-2000 (first entry)  
 DT Streptococcus pyogenes GRAB peptide EKL 24.  
 XX  
 DE GRAB: protein G related alpha2M binding protein; vaccine;  
 KW alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;  
 KM immune response; Streptococcus pyogenes infection; peptide EKL 24.  
 XX  
 OS Streptococcus pyogenes.  
 XX  
 PN WO200026240-A2.  
 XX  
 PD 11-MAY-2000.  
 XX  
 PF 02-NOV-1999; 99WO-GB03631.  
 XX  
 PR 02-NOV-1998; 98GB-0023975.  
 XX  
 PA (ACTI-) ACTINOVA LTD.  
 XX  
 PI Bjorck LH, Rasmussen M;  
 DR WPI: 2000-365572/31.  
 XX  
 PT New alpha2M binding protein for generating a protective immune response  
 PT to group A streptococcus and purifying the binding protein -  
 XX  
 PS Example 9; Page 31; 67pp; English.  
 XX  
 CC The patent discloses a new family of proteins termed GRAB (protein G  
 CC related alpha2M binding protein) from Streptococcus pyogenes which have  
 CC the ability to bind alpha2-macroglobulin (alpha2M) and show homology to  
 CC protein G of group G Streptococcus. GRAB protein and peptides derived  
 CC from it are used in vaccine compositions for generating a protective  
 CC immune response against group A Streptococcus. Antibodies against GRAB  
 CC are useful for treating Streptococcus pyogenes infections. The present  
 CC is also useful for purifying alpha2M from a sample. The present sequence  
 CC is a GRAB peptide EKL 24 useful in vaccine composition. It was used  
 CC to produce immunoglobulin G (IgG) antibodies specific for native GRAB  
 CC protein in sheep. The peptide has a cysteine insert at the C-terminal  
 CC for attachment to a hetero-bifunctional linker.  
 CC  
 XX  
 SQ Sequence 25 AA:  
 Query Match 14.8%; Score 115; DB 21; Length 25;  
 Best Local Similarity 100.0%; Pred. No. 0.00091;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 26 EKILRNERRAIDELKKOAIEDKE 49  
 DB 1 EKILRNERRAIDELKKOAIEDKE 24  
 RESULT 27  
 AAR71928  
 ID AAR71928 standard; Protein; 413 AA.  
 AC AAR71928;  
 DT 22-SEP-1995 (first entry)  
 XX  
 DE S. dysgalactiae MAG protein.  
 XX  
 KW MAG; fast alpha-2-macroglobulin binding protein; FAM;  
 KM plasma proteinase-inhibitor binding protein.  
 XX  
 OS Streptococcus dysgalactiae strain 8215.  
 XX  
 FH Key Location/Qualifiers  
 FT Peptide 1..34  
 FT /label= sig.peptide  
 FT Domain 35..92

FT /label= A2M  
 FT /note= "alpha-2-macroglobulin binding domain"  
 FT Domain 93..242  
 FT /label= Alb  
 FT /note= "albumin binding domain"  
 FT Domain 243..312  
 FT /label= IgG  
 FT /note= "IgG binding domain"  
 FT Region 313..387  
 FT /note= "cell wall region"  
 FT Region 388..413  
 FT /label= membrane-spanning region  
 PN WO9507296-A.  
 XX  
 PD 16-MAR-1995.  
 XX  
 PF 06-SEP-1994; 94WO-SE00826.  
 XX  
 PR 06-SEP-1993; 93SE-0002855.  
 XX  
 PA (GUS/) GUS B.  
 PA (JONS/) JONSSON H.  
 PA (LIND/) LINDBERG M.  
 PA (MUEL/) MUELLER H.  
 PA (RANT/) RANTAMAKI L K.  
 XX  
 PI Guss B, Jonsson H, Lindberg M, Mueller H, Rantamaki LK;  
 XX  
 DR WPI: 1995-123382/16.  
 DR N-PSDB: AAQ89197.  
 XX  
 PT DNA encoding fast alpha 2-macroglobulin-binding proteins - used  
 PT to obtain prods. for sepn., detection or quantification or for  
 PT binding inhibition  
 PS Disclosure; Fig.4; 50pp; English.  
 XX  
 CC A phage lambda EMBL library of S. dysgalactiae 8215 DNA was analyzed  
 CC for fast-acting alpha-2-macroglobulin (FAM)-, BSA- and IgG-binding  
 CC activity. Clone lambda-SD1, which expressed all 3 activities, was  
 CC analyzed to obtain DNA encoding the FAM-binding protein, MAG. The  
 CC mag gene (given in Q89197) encodes a 44 kDa protein (R71929).  
 CC  
 XX  
 SQ Sequence 413 AA:  
 Query Match 14.7%; Score 114.5; DB 16; Length 413;  
 Best Local Similarity 26.2%; Pred. No. 0.033;  
 Matches 49; Conservative 15; Mismatches 56; Indels 67; Gaps 6;  
 OY 36 AIDELKKOAIED-----KEATTAIE---AASDLEALADQTDALQSEAAVVK 81  
 DB 206 AIEILKKRYIGDYIKLINNGKTAEGYALNDEILASPAVIDAELTPALTYTKLYI-- 263  
 OY 82 ADNAAADALEALADQTDALQSEEA-EVYQSDNAAASDAVE----- 119  
 DB 264 --NGKTLKGETTTAKVADAEAKAFKQYANANGVDGVVYTDATKRTFTVEMTEVPDCA 321  
 OY 120 -----KAAPIALDVKKTKDT-----KPVYKKEERQNVNLTPT 152  
 DB 322 PTERPKKPEASIFLVLPFATPIAKDKDKDTKKDKKEDAKKPEAKKPEAKKAATLPT 381  
 OY 153 TGEESNP 159  
 DB 382 TGEESNP 388  
 RESULT 28  
 AAR71670  
 ID AAR71670 standard; Protein; 413 AA.  
 AC AAR71670;

```
DT 21-SEP-1995 (first entry)
XX
DE S. dysgalactiae MAG protein.
XX
XX Mag gene; serum albumin binding protein; protein stabilisation;
KW vaccine.
XX
OS Streptococcus dysgalactiae strain 8215.
XX
XX Key Location/Qualifiers
FH 1..34
FT Peptide /label= Sig.peptide
FT 35..92
FT Domain /label= A2-M
FT /note= "alpha-2-macroglobulin binding domain"
FT 93..242
FT Domain /label= Alb
FT /note= "albumin binding domain"
FT 243..312
FT /label= IgG
FT /note= "IgG binding domain"
FT 313..387
FT Region /note= "cell wall binding region"
FT 388..413
FT Region /note= "membrane-spanning region"
FT
XX
XX WO9507300-A.
XX
XX 16-MAR-1995.
XX
XX 06-SEP-1994; 94WO-SE00825.
XX
XX 06-SEP-1993; 93SE-0002856.
XX
XX (GUS/) GUS B.
PA (JONS/) JONSSON H.
PA (LIND/) LINDBERG M.
XX
XX Guss B, Jonsson H, Lindberg M;
XX
XX WPI: 1995-123386/16.
DR N-PSDB: AAO86080.
XX
XX DNA encoding a serum albumin binding protein - used to obtain
PT prods., detection, quantification, protein
PT stabilisation or vaccine development
XX
XX Disclosure: Page 29-30; 38pp; English.
XX
XX A phage lambda EMBL3 library of S. dysgalactiae 8215 DNA was
CC analyzed for alpha-2-macroglobulin-, BSA- and IGG-binding activity.
CC Clone lambda-SD1, which expressed all 3 activities, was analyzed to
CC obtain DNA encoding the SA-binding protein, MAG. The mag gene
CC (given in Q86080) encoded a 44 kDa MAG protein (R71670).
CC Recombinant, immobilized MAG was used for serum albumin affinity
CC purification, detection and assay.
XX
XX Sequence 413 AA:
SQ
Query Match 14.7%; Score 114.5; DB 16; Length 413;
Best Local Similarity 26.2%; Pred. NO. 0.033;
Matches 49; Conservative 15; Mismatches 56; Indels 67; Gaps 6;
```

```
DB 322 PTEPKKPEASIPVLPATPIAKDDAKKDDTKKDDTKREDAKKPEAKKEAKKATLPT 381
OY 153 TGEESNP 159
DB 382 TGEESNP 388
RESULT 29
AAG38476
ID AAG38476 standard; Protein: 154 AA.
XX
XX AAG38476;
AC
XX
XX 18-OCT-2000 (first entry)
DE Arabidopsis thaliana protein fragment SEQ ID NO: 47473.
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
XX Arabidopsis thaliana.
PN
XX
XX EPI033405-A2.
XX
XX 06-SEP-2000.
PD
XX
XX 25-FEB-2000; 2000EP-0301439.
PF
XX
XX 25-FEB-1999; 99US-0121825.
XX
XX 05-MAR-1999; 99US-0123180.
XX
XX 09-MAR-1999; 99US-0123548.
XX
XX 23-MAR-1999; 99US-0125788.
XX
XX 25-MAR-1999; 99US-0126264.
XX
XX 29-MAR-1999; 99US-0126785.
XX
XX 01-APR-1999; 99US-0127462.
XX
XX 06-APR-1999; 99US-0128234.
XX
XX 08-APR-1999; 99US-0128714.
XX
XX 16-APR-1999; 99US-0129845.
XX
XX 19-APR-1999; 99US-0130077.
XX
XX 21-APR-1999; 99US-0130449.
XX
XX 23-APR-1999; 99US-0130510.
XX
XX 28-APR-1999; 99US-0130891.
XX
XX 30-APR-1999; 99US-0131449.
XX
XX 30-APR-1999; 99US-0132048.
XX
XX 04-MAY-1999; 99US-0132407.
XX
XX 05-MAY-1999; 99US-0132484.
XX
XX 06-MAY-1999; 99US-0132485.
XX
XX 06-MAY-1999; 99US-0132486.
XX
XX 07-MAY-1999; 99US-0132487.
XX
XX 11-MAY-1999; 99US-0132853.
XX
XX 14-MAY-1999; 99US-0134218.
XX
XX 14-MAY-1999; 99US-0134219.
XX
XX 14-MAY-1999; 99US-0134221.
XX
XX 14-MAY-1999; 99US-0134370.
XX
XX 18-MAY-1999; 99US-0134768.
XX
XX 19-MAY-1999; 99US-0134941.
XX
XX 20-MAY-1999; 99US-0135124.
XX
XX 21-MAY-1999; 99US-0135353.
XX
XX 24-MAY-1999; 99US-0135629.
XX
XX 25-MAY-1999; 99US-0136021.
XX
XX 27-MAY-1999; 99US-0136392.
XX
XX 28-MAY-1999; 99US-0136782.
XX
XX 01-JUN-1999; 99US-0137222.
XX
XX 03-JUN-1999; 99US-0137528.
XX
XX 04-JUN-1999; 99US-0137502.
XX
XX 07-JUN-1999; 99US-0137724.
XX
XX 08-JUN-1999; 99US-0138094.
XX
XX 10-JUN-1999; 99US-0138540.
XX
XX 10-JUN-1999; 99US-0138847.
XX
XX 14-JUN-1999; 99US-0139119.
XX
XX 16-JUN-1999; 99US-0139452.
```



PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139889.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144652.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 26-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145921.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.

PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149920.  
PR 23-AUG-1999; 99US-0149920.  
PR 23-AUG-1999; 99US-0149920.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 14.4%; Score 112; DB 21; Length 154;  
Best Local Similarity 26.5%; Pred. No. 0.016;  
Matches 40; Conservative 25; Mismatches 58; Indels 28; Gaps 5;

Oy 17 LTNLLGNAP-----EKLRLNDEERAIIDLKQAIEDKATTAIFAASSDALEALADOTDA 71  
Db 1 MONEGGSVPNKRPESENVVAKENNTSEGEKONQTAETRETSVEAKETFPPEPKETTPA 60  
Oy 72 LQSE-----BAAYV---KADNAASDALEALADOTDALQSEAEAVQSDNAAS 115

```
DB      61 VOPEVAAVESSADAGEAAVAAPKEVENAATENAEKAEVAVAAPKEVEY-----AV 114
OY      116 DAMEKA-ATP1ALDVKKTQTKRPVVKKEBRO 145
        : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      115 EAKKKAEPVKAEPVKAEPVKAEPVKEESKQ 145

RESULT 30
AAP70468
ID      AAP70468 standard; Protein: 480 AA.
XX
AC      AAP70468;
XX
DT      21-MAY-1991 (first entry)
XX
DE      Sequence of polypeptide possessing IgG-binding activity of
DE      protein G from Streptococcus G148.
XX
KW      Antibody-binding; IgG; IgA; Immunoglobulin.
XX
OS      Streptococcus G148.
XX
PN      WO8705631-A.
XX
PD      24-SEP-1987.
XX
PF      20-MAR-1987; 87WO-SE00145.
XX
PR      21-MAR-1986; 86SE-0001325.
XX
PA      (PHAA ) PHARMACIA AB.
XX      (GUS5/) GUS5 B M.
XX
PI      Guse BM, Lindberg KM, Flock JI, Uhlen CEM;
XX
DR      WPT: 1987-277686/39.
DR      N-PSDB; AAN70757.
XX
PT      New recombinant DNA molecules - for producing proteins with
PT      IgG-binding specificity of protein G or proteins A and G
XX
PS      Example: Fig 2: 39pp; English.
XX
CC      A recombinant DNA molecule containing a nucleotide sequence which
CC      codes for a protein or polypeptide having the same IgG specificity as
CC      protein G from Streptococcus G148 (AAN70757) is claimed. See, for
CC      example, AAN70754, AAN70755 and AAN70756.
XX
SQ      Sequence 480 AA;

Query Match      14.4%; Score 112; DB 8; Length 480;
Best Local Similarity 31.0%; Pred. No. 0.066;
Matches 44; Conservative 12; Mismatches 62; Indels 24; Gaps 5;

OY      41 KKAQIEDKEATTAIE---AASSDAL--EALADQTDALQSEEAAYVKAADMAADALEALA 94
        : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      315 KEPEVIDASELTPAVTVYKLVNGTKLKGEBTTKAVDAEFAEKAQYANDNGVWTVYD 374
        : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY      95 DQTDALQSEEAAYVOSDNMAADAMER-----AATP1ALDVKKTQTKD-----KP 137
        : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      375 DATKTFVTTEM-VTEVPDGAFTPEPEKPEASIPLVPLTPATPIAKDDAKKDDTKKEDAKRP 433
        : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY      138 VVKKERONVNTLPTTGEEBSP 159
        : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      434 EAKKEDAKKAETLPTTGEEBSP 455

RESULT 31
AAG41392
ID      AAG41392 standard; Protein: 191 AA.
XX
AC      AAG41392;
```

```
XX      18-OCT-2000 (first entry)
XX
DT      Arabidopsis thaliana protein fragment SEQ ID NO: 51493.
XX
DE      Protein identification; signal transduction pathway; metabolic pathway;
XX      hybridisation assay; genetic mapping; gene expression control; promoter;
XX      termination sequence.
XX
OS      Arabidopsis thaliana.
XX
PN      EP1033405-A2.
XX
PD      06-SEP-2000.
XX
PF      25-FEB-2000; 2000EP-0301439.
XX
PR      25-FEB-1999; 99US-0121825.
PR      05-MAR-1999; 99US-0123180.
PR      09-MAR-1999; 99US-0123548.
PR      23-MAR-1999; 99US-0125788.
PR      25-MAR-1999; 99US-0126264.
PR      29-MAR-1999; 99US-0126785.
PR      01-APR-1999; 99US-0127462.
PR      06-APR-1999; 99US-0128234.
PR      08-APR-1999; 99US-0128714.
PR      16-APR-1999; 99US-0129845.
PR      19-APR-1999; 99US-0130077.
PR      21-APR-1999; 99US-0130449.
PR      23-APR-1999; 99US-0130510.
PR      28-APR-1999; 99US-0130891.
PR      30-APR-1999; 99US-0131449.
PR      30-APR-1999; 99US-0132048.
PR      04-MAY-1999; 99US-0132407.
PR      05-MAY-1999; 99US-0132484.
PR      06-MAY-1999; 99US-0132485.
PR      06-MAY-1999; 99US-0132486.
PR      07-MAY-1999; 99US-0132487.
PR      11-MAY-1999; 99US-0132863.
PR      14-MAY-1999; 99US-0134218.
PR      14-MAY-1999; 99US-0134219.
PR      14-MAY-1999; 99US-0134221.
PR      18-MAY-1999; 99US-0134370.
PR      19-MAY-1999; 99US-0134768.
PR      20-MAY-1999; 99US-0134941.
PR      21-MAY-1999; 99US-0135124.
PR      24-MAY-1999; 99US-0135353.
PR      24-MAY-1999; 99US-0135629.
PR      25-MAY-1999; 99US-0136021.
PR      27-MAY-1999; 99US-0136392.
PR      28-MAY-1999; 99US-0136782.
PR      01-JUN-1999; 99US-0137222.
PR      03-JUN-1999; 99US-0137528.
PR      04-JUN-1999; 99US-0137502.
PR      07-JUN-1999; 99US-0137724.
PR      08-JUN-1999; 99US-0138094.
PR      10-JUN-1999; 99US-0138540.
PR      10-JUN-1999; 99US-0138847.
PR      14-JUN-1999; 99US-0139119.
PR      16-JUN-1999; 99US-0139452.
PR      17-JUN-1999; 99US-0139453.
PR      18-JUN-1999; 99US-0139492.
PR      18-JUN-1999; 99US-0139454.
PR      18-JUN-1999; 99US-0139455.
PR      18-JUN-1999; 99US-0139456.
PR      18-JUN-1999; 99US-0139457.
PR      18-JUN-1999; 99US-0139458.
PR      18-JUN-1999; 99US-0139459.
PR      18-JUN-1999; 99US-0139460.
PR      18-JUN-1999; 99US-0139461.
PR      18-JUN-1999; 99US-0139462.
PR      18-JUN-1999; 99US-0139463.
```



AC AAC1391;  
XX 18-OCT-2000 (first entry)  
XX  
XX Arabidopsis thaliana protein fragment SEQ ID NO: 51492.  
DE  
XX Protein identification; signal transduction pathway; metabolic pathway;  
KM hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
XX EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
XX 25-FEB-2000; 2000EP-0301439.  
XX  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 03-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 23-AUG-1999; 99US-0150566.

PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 14.0%; Score 109; DB 21; Length 230;

Best Local Similarity 27.3%; Pred. No. 0.049; Mismatches 51; Indels 56; Gaps 7;

QY 27 KLAIRNEBRAIDELK-----QATIED-----KEATTAIEAASDLEALADOTDALQ 73  
DB 46 KLEIOWNEKASELKEIQRERDHOATIEDLKKEKTKDAKTQLSLEELKIAVFEAQEKAD 105  
QY 74 SEE-----AAVYKADNA-----ASDALEALADOT-DALOSEEAFAVQSD 111  
DB 106 AEEHARERLNVAVLESDFRSIAVYKESAELTETEARACRDETLKTIEMSERETEDIK 165  
QY 112 NAASDAWEKA-----ATPIALDVKK-----TKDTKPYVKKKEERON 146  
DB 166 AATQDALKKAMQAEATIVVDELKRRRKAASRIIAESKMAKSKTKEVLSKSPRS 221

RESULT 33  
AAG41390  
ID AAG41390 standard; Protein; 316 AA.

XX AC AAG41390;  
XX DT 18-OCT-2000 (first entry)  
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 51491.  
XX KW Protein identification; signal transduction pathway; metabolic pathway;  
XX KW hybridisation assay; genetic mapping; gene expression control; promoter;  
XX KW termination sequence.  
XX OS Arabidopsis thaliana.  
XX PN EP1033405-A2.  
XX PD 06-SEP-2000.  
XX PF 25-FEB-2000; 2000EP-0301439.  
XX PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130077.  
PR 23-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132407.  
PR 05-MAY-1999; 99US-0132484.  
PR 06-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139453.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.

PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140685.  
PR 28-JUN-1999; 99US-0140823.  
PR 30-JUN-1999; 99US-0140991.  
PR 01-JUL-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142255.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 19-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 20-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 21-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 23-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145813.  
PR 27-JUL-1999; 99US-0145818.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 10-AUG-1999; 99US-0147935.  
PR 11-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0148368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0148426.  
PR 20-AUG-1999; 99US-0148722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.

PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151085.  
PR 27-AUG-1999; 99US-0151086.  
PR 27-AUG-1999; 99US-0151087.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155559.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159337.  
PR 18-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161922.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 14.0%; Score 109; DB 21; Length 316;  
Best Local Similarity 27.3%; Pred. No. 0.072;  
Matches 48; Conservative 21; Mismatches 51; Indels 56; Gaps 7;

QY 27 KALRNEERAIIDELKK-----QAIED-----KATTAIEAASDLEALADODIALO 73  
DB 132 KLELWVERKAHSELKELEBORERDHOAIEDLKKEFKAKTQULSLEELKIAVEAQAED 191  
QY 74 SEE-----AAVKKADNA-----ASDALEALADOT-DALOSEAEAYQSD 111  
DB 192 AEEHARERLNVAVLESDFSLAVYKESABELTETALACRDETIKTELEMSREIEDIK 251  
QY 112 NAASDAWEKA-----ATPIALDYVK-----TKOTKPYVKKKEEON 146  
DB 252 AATODALKRAEMAOEATVIYVDLELKRKRKAASRIILAESKWKCAKSTFEVLKSKPRSS 307

RESULT 34  
AAG20162

ID AAG20162 standard; Protein: 180 AA.  
XX  
AC AAG20162;  
XX  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 22242.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
PN EPI033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123160.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125768.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148317.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.

PR	23-AUG-1999;	99US-0149930.
PR	25-AUG-1999;	99US-0150566.
PR	26-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151066.
PR	27-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151438.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.
PR	13-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.

	Query Match	14.0%	Score 108.5	DB 21	Length 180	
	Best Local Similarity	24.0%	Pred. No. 0.04	74	Indels 9	Gaps 4
	Matches	36	Conservative	31	Mismatches	7
Oy	12 PNCGLTLLCNGAP-----EKIALNBERAIDELKQATEDEKAATTAIPAASDALLEALA	66				
	:   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :					
Db	8 PKESYMQHEEGSVPNPKPVSENVVAENNTTESGEKONQTVAEETSTVSVEAKETTSVEPTPK	67				
Oy	67 DQTDLQOSEEAAYVADNAASDALPALA-DQTDLQOSEEAAYQSDNNAADMKEKAATPI	125				
	:   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :					
Db	68 ETTPAAREVAAVESSSAAGGAAGAAYAPBEKERKATENAEE-AKKEAVAVAAPER--VEY	124				
Oy	126 ALDYKKTKDTKPYVKKKERONNTLPTTGG	155				
	: :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :					
Db	125 AVEAEKKAEAEPPVKEAEPPVKAEBEPVKAEE	154				

AB62318	ABB62318 standard; Protein; 1096 AA.
AC	ABB62318;
AD	26-MAR-2002 (first entry)
AE	Drosophila melanogaster polypeptide SEQ ID NO 13746.
AF	Drosophila; developmental biology; cell signalling; insecticide;
AG	pharmaceutical.
AH	Drosophila melanogaster.
AI	WO200171042-A2.
AJ	27-SEP-2001.
AK	23-MAR-2001; 2001WO-US09231.
AL	23-MAR-2000; 2000US-191637P.
AM	11-JUL-2000; 2000US-0614150.
AN	(PEKE ) PE CORP NY.
AO	Venter JC, Adams M, Li PWD, Myers EW;
AP	WPI; 2001-656860/75.
AQ	N-PSDB; ABL06421.
AR	New isolated nucleic acid detection reagent for detecting 1000 or more
AS	genes from Drosophila and for elucidating cell signalling and cell-cell
AT	interactions -
AW	Disclosure; SEQ ID NO 13746; 21pp + Sequence Listing; English.
AX	The invention relates to an isolated nucleic acid detection reagent
AY	capable of detecting 1000 or more genes from Drosophila. The invention is
AZ	useful in developmental biology and in elucidating cell signalling and
BA	cell-cell interactions in higher eukaryotes for the development of
BB	insecticides, therapeutics and pharmaceutical drugs. The invention
BC	discloses genomic DNA sequences (ABL6176-ABL30511), expressed DNA
BD	sequences (ABL01840-ABL16175) and the encoded proteins
BE	(ABH57737-ABH72002).
BF	The sequence data for this patent did not form part of the printed
BG	specification, but was obtained in electronic format directly from WIPO
BH	at ftp.wipo.int/pub/published_pct_sequences.
BI	Sequence 1096 AA;

[illegible]



DT 17-OCT-2000 (first entry)  
XX Arabidopsis thaliana protein fragment SEQ ID NO: 22243.  
DE  
XX Protein Identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
XX termination sequence.  
OS Arabidopsis thaliana.  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000: 2000EP-0301439.  
XX  
XX 25-FEB-1999: 99US-0121825.  
PR 05-MAR-1999: 99US-0123180.  
PR 09-MAR-1999: 99US-0123548.  
PR 23-MAR-1999: 99US-0125788.  
PR 25-MAR-1999: 99US-0126264.  
PR 29-MAR-1999: 99US-0126785.  
PR 01-APR-1999: 99US-0127462.  
PR 06-APR-1999: 99US-0128234.  
PR 08-APR-1999: 99US-0128714.  
PR 16-APR-1999: 99US-0129845.  
PR 19-APR-1999: 99US-0130077.  
PR 21-APR-1999: 99US-0130449.  
PR 23-APR-1999: 99US-0130510.  
PR 23-APR-1999: 99US-0130891.  
PR 28-APR-1999: 99US-0131449.  
PR 30-APR-1999: 99US-0132048.  
PR 30-APR-1999: 99US-0132407.  
PR 04-MAY-1999: 99US-0132484.  
PR 05-MAY-1999: 99US-0132485.  
PR 06-MAY-1999: 99US-0132486.  
PR 07-MAY-1999: 99US-0132487.  
PR 11-MAY-1999: 99US-0132863.  
PR 14-MAY-1999: 99US-0134256.  
PR 14-MAY-1999: 99US-0134218.  
PR 14-MAY-1999: 99US-0134219.  
PR 14-MAY-1999: 99US-0134221.  
PR 14-MAY-1999: 99US-0134370.  
PR 18-MAY-1999: 99US-0134768.  
PR 19-MAY-1999: 99US-0134841.  
PR 21-MAY-1999: 99US-0135124.  
PR 21-MAY-1999: 99US-0135153.  
PR 24-MAY-1999: 99US-0135629.  
PR 25-MAY-1999: 99US-0136021.  
PR 27-MAY-1999: 99US-0136392.  
PR 28-MAY-1999: 99US-0136782.  
PR 01-JUN-1999: 99US-0137222.  
PR 03-JUN-1999: 99US-0137528.  
PR 04-JUN-1999: 99US-0137502.  
PR 07-JUN-1999: 99US-0137724.  
PR 08-JUN-1999: 99US-0138094.  
PR 10-JUN-1999: 99US-0138440.  
PR 10-JUN-1999: 99US-0138847.  
PR 14-JUN-1999: 99US-0139119.  
PR 16-JUN-1999: 99US-0139452.  
PR 17-JUN-1999: 99US-0139453.  
PR 17-JUN-1999: 99US-0139492.  
PR 18-JUN-1999: 99US-0139454.  
PR 18-JUN-1999: 99US-0139455.  
PR 18-JUN-1999: 99US-0139456.  
PR 18-JUN-1999: 99US-0139457.  
PR 18-JUN-1999: 99US-0139458.  
PR 18-JUN-1999: 99US-0139459.  
PR 18-JUN-1999: 99US-0139460.  
PR 18-JUN-1999: 99US-0139461.  
PR 18-JUN-1999: 99US-0139462.  
PR 18-JUN-1999: 99US-0139463.  
PR 18-JUN-1999: 99US-0139750.  
  
PR 18-JUN-1999: 99US-0139763.  
PR 21-JUN-1999: 99US-0139817.  
PR 22-JUN-1999: 99US-0139889.  
PR 23-JUN-1999: 99US-0140353.  
PR 24-JUN-1999: 99US-0140354.  
PR 24-JUN-1999: 99US-0140695.  
PR 28-JUN-1999: 99US-0140823.  
PR 29-JUN-1999: 99US-0140991.  
PR 30-JUN-1999: 99US-0141287.  
PR 01-JUL-1999: 99US-0141842.  
PR 01-JUL-1999: 99US-0142154.  
PR 02-JUL-1999: 99US-0142055.  
PR 06-JUL-1999: 99US-0142390.  
PR 08-JUL-1999: 99US-0142803.  
PR 09-JUL-1999: 99US-0142820.  
PR 12-JUL-1999: 99US-0142977.  
PR 13-JUL-1999: 99US-0143542.  
PR 14-JUL-1999: 99US-0143624.  
PR 15-JUL-1999: 99US-0144005.  
PR 16-JUL-1999: 99US-0144085.  
PR 16-JUL-1999: 99US-0144086.  
PR 19-JUL-1999: 99US-0144325.  
PR 19-JUL-1999: 99US-0144331.  
PR 19-JUL-1999: 99US-0144332.  
PR 19-JUL-1999: 99US-0144333.  
PR 19-JUL-1999: 99US-0144334.  
PR 19-JUL-1999: 99US-0144335.  
PR 20-JUL-1999: 99US-0144352.  
PR 20-JUL-1999: 99US-0144632.  
PR 20-JUL-1999: 99US-0144684.  
PR 21-JUL-1999: 99US-0144814.  
PR 21-JUL-1999: 99US-0145086.  
PR 21-JUL-1999: 99US-0145088.  
PR 22-JUL-1999: 99US-0145085.  
PR 22-JUL-1999: 99US-0145087.  
PR 22-JUL-1999: 99US-0145089.  
PR 22-JUL-1999: 99US-0145192.  
PR 23-JUL-1999: 99US-0145145.  
PR 23-JUL-1999: 99US-0145218.  
PR 23-JUL-1999: 99US-0145224.  
PR 26-JUL-1999: 99US-0145276.  
PR 27-JUL-1999: 99US-0145913.  
PR 27-JUL-1999: 99US-0145918.  
PR 27-JUL-1999: 99US-0145919.  
PR 28-JUL-1999: 99US-0145951.  
PR 02-AUG-1999: 99US-0146386.  
PR 02-AUG-1999: 99US-0146388.  
PR 03-AUG-1999: 99US-0146389.  
PR 03-AUG-1999: 99US-0147038.  
PR 04-AUG-1999: 99US-0147204.  
PR 04-AUG-1999: 99US-0147302.  
PR 05-AUG-1999: 99US-0147192.  
PR 05-AUG-1999: 99US-0147260.  
PR 06-AUG-1999: 99US-0147303.  
PR 06-AUG-1999: 99US-0147416.  
PR 09-AUG-1999: 99US-0147493.  
PR 09-AUG-1999: 99US-0147935.  
PR 10-AUG-1999: 99US-0148171.  
PR 11-AUG-1999: 99US-0148319.  
PR 12-AUG-1999: 99US-0148341.  
PR 13-AUG-1999: 99US-0148565.  
PR 13-AUG-1999: 99US-0148684.  
PR 16-AUG-1999: 99US-0149368.  
PR 17-AUG-1999: 99US-0149175.  
PR 18-AUG-1999: 99US-0149426.  
PR 20-AUG-1999: 99US-0149722.  
PR 20-AUG-1999: 99US-0149723.  
PR 20-AUG-1999: 99US-0149929.  
PR 23-AUG-1999: 99US-0149902.  
PR 23-AUG-1999: 99US-0149930.  
PR 25-AUG-1999: 99US-0150566.  
PR 26-AUG-1999: 99US-0150884.  
PR 27-AUG-1999: 99US-0151065.

PR	27-AUG-1999;	99US-0151066.
PR	27-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151438.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.
PR	13-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	23-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.

Query Match	13.6%	Score 105.5	DB 21	Length 168
Best Local Similarity	24.1%	Pred. No. 0.068		
Matches	35	Conservative	31	Mismatches 70; Indels 9; Gaps 4
QY	17	LTNLLGNAP----	EKLALRNERAIDELKKAQIDKEATTAIEAASSDALEALADQDA	71
Db	1	MONEGSAVPNKPVSSENNVAKENNTESGEGONQTVAFSTTSVEAKETSPVEPTKETTPA		60
QY	72	LOSEEAAYVKADNAASDALEALA-DOTALOSEEAEEVYQSDNAASDAWEKAATPIALDVK		130
Db	61	AEPEVAAYVESSSSAAGEAANAAPKPEVKAATENAE-AKVEAIVAAPAPEK--VEVAVEAE		117
QY	131	KTKQTKPVYKKKEERONVNTLPPTGE		155
Db	118	KKAEEAPVKVEAEPPVKAEEAPVKAEE		142
RESULT 37				
AAW27286				
ID	AAW27286	standard	Protein	314 AA.
XX				
AC	AAW27286			

xx	27-APR-1998	(first entry)
DT		
xx		
DE		
xx		
xx		
KW	EsPB; rabbit enteropathogenic E. coli; RDEC-1; virulence factor;	
KW	pathogen; infection; diagnosis; therapy; vaccine; gastroenteritis.	
xx		
OS	Escherichia coli.	
xx		
PN	WO9740063-A2.	
xx		
PD	30-OCT-1997.	
xx		
xx		
PF	23-APR-1997; 97WO-CA00265.	
xx		
PR	23-APR-1996; 96US-0015999.	
xx		
xx		
PA	(UYBR-) UNIV BRITISH COLUMBIA.	
xx		
PI	Finlay BB, Kenny B, Stein M;	
xx		
DR	WPL 1997-535772/49.	
DR	N-PSDB; AAT91243.	
xx		
PT	EsPB from entero-pathogenic or entero-haemorrhagic E. coli - used to	
PT	immunise against, and treat diseases caused by EsPB producing	
PT	organisms	
xx		
PS	Example 6; Fig 4B; 62pp; English.	
xx		
CC	This sequence comprises the EsPB virulence factor of a rabbit	
CC	enteropathogenic Escherichia coli (RDEC-1) strain. Its amino acid	
CC	sequence was deduced from an isolated espb gene (see AAT91293). The	
CC	predicted EsPB polypeptide shows sequence homology more to	
CC	enterohaemorrhagic E. coli (EHEC) than to enteropathogenic E. coli	
CC	(EHEC). Claimed novel, isolated EsPB polypeptides (see also	
CC	AAW7284-5) are characterised by: (a) being a secreted protein from	
CC	EPEC or EHEC; and (b) having a mol.wt. of about 25 kD as determined	
CC	by SDS-PAGE. Recombinant EsPB polypeptides can be expressed in host	
CC	cells. They can be used to immunise a host susceptible to disease	
CC	caused by EsPB-producing E. coli, and in a method for ameliorating	
CC	disease caused by EsPB-producing organisms. Antibodies raised	
CC	against EsPB can be used in the diagnosis of infection.	
xx		
Sequence	314 AA;	
xx		

Query Match	13.6%	Score 105.5	DB 18	Length 314
Best Local Similarity	26.7%	Pred. No. 0.15		
Matches 45	Conservative 23	Mismatches 68	Indels 21	Gaps 6
QY	14	GCTTLNLGNAPKELALNEERATIDELKQA--IEDKEAT--AIENASSDALEALADQ	68	
		:::	:::	:
Db	100	GGAISVVG-----ILGSFAIINSATKSGASDIAGKTASTSSKALIDAASTATATKTLTKA	152	
		:::	:::	:
QY	69	T---DALQSEAAVVKADNNAASDALEALEADQTDALQSEAEVQSDNNAASDAANEKATATP	124	
		:::	:::	:
Db	153	TESVADAEVDASSVQQAATTATRAASTSYADVIDADIASQARASQLAENNADAAQKASRA	212	
		:::	:::	:
QY	125	---IALDVKKTKDKPKPVKKKEERONVTLPPTGSES	157	
		:::	:::	:
Db	213	SRFMAAYD-KITGSTPIAVTSLAEGTKTLPTTVSES	248	
		:::	:::	:
RESULT 38				
ID	AAK71127	standard; Protein; 208 AA.		
xx	AAK71127;			
AC	AAK71127;			
xx				
DT	11-OCT-1995	(first entry)		
xx				
DE	Synthetic protein G from gene MD1.			

```

XX Synthetic: protein G; IgG; Fc receptors; heavy chain; leukaemia;
KM Lymphoma; cancer; autoimmune disease.
XX OS Synthetic.
XX PN W09506125-A.
XX PD 02-MAR-1995.
XX PF 23-AUG-1994; 94WO-US09141.
XX PR 23-AUG-1993; 93US-0110653.
XX PA (IMM-) APPLIED IMMUNE SCI INC.
XX PI Lee YM, Okarma TB, Talib S;
XX DR WPI: 1995-106854/14.
XX DR N-PSDB: AA084886.
XX PT New peptide(s) which bind the Fc region of an immunoglobulin -
XX PT comprising a nontotal portion of the amino acid sequence of
XX PT Protein A and/or Protein G
XX PS Disclosure: Fig 4A; 61pp; English.
XX CC The sequence is the prod. of the synthetic protein G gene, MD-1. The
XX CC proteins produced by MD-1 are capable of specifically binding
XX CC the constant region of the heavy chain of IgG in the same way as
XX CC neutral Fc receptors. They can be used to analyse the structure and
XX CC function of Fc receptors, as well as in antibody production, cell
XX CC culture, diagnosis and therapy. They can be used to treat disorders
XX CC such as leukaemia and lymphoma, cancer and immune disorders.
XX CC See also AAR71123-8.
XX SQ Sequence 208 AA:

Query Match 13.5%; Score 105; DB 16; Length 208;
Best Local Similarity 29.8%; Pred. No. 0.098;
Matches 42; Conservative 12; Mismatches 63; Indels 24; Gaps 4;

QY 41 KKOAIEDKERTATIEA-----ASSDALFALADOTDLOSEEAAYVKADNAASDALEALA 94
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 62 KPEVIDASELTTPAVTYTKLYIGKTLKGETTTRAVDAETAEKAFKOYANOGVGVTYD 121
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 95 DQTDALQSEAEVYQSDNAASDAMEK-----AATPIADVKTKDT-----KP 137
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 122 DATKTFVTTEM-VTEVVGADAPTEPEKEASIPVPLTPATPIAKDKAKKDDTKKEDAKKP 180
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 138 VVKKEKRONVNTLPTTGEESEN 158
   ||| : ||| ||| ||| |||
DB 181 EAKKEDAKKAEKTLPTTGEGSN 201
   ||| : ||| ||| ||| |||

RESULT 39
AAR71929
ID AAR71929 standard; Protein: 664 AA.
XX
AC AAR71929;
XX
DT 22-SEP-1995 (first entry)
XX
DE S. dysgalactiae MIG.
XX
KM MIG: fast alpha-2-macroglobulin binding protein; FAM;
KM Plasma proteinase-inhibitor binding protein.
XX
OS Streptococcus dysgalactiae strain SCL.
FH Key Location/Qualifiers
FT Peptide 1..30
FT /label= Sig_peptide

```

```

FT Domain 209..278
FT /label= IgG1
FT /note= "IgG binding domain 1"
FT Domain 279..348
FT /label= IgG2
FT /note= "IgG binding domain 2"
FT Domain 349..418
FT /label= IgG3
FT /note= "IgG binding domain 3"
FT Domain 419..488
FT /label= IgG4
FT /note= "IgG binding domain 4"
FT Domain 489..558
FT /label= IgG5
FT /note= "IgG binding domain 5"
FT Region 559..638
FT /note= "cell wall spanning region"
FT Peptide 630..635
FT /note= "putative wall anchoring motif"
FT Region 639..664
FT /note= "membrane spanning domain"

W09507296-A.
16-MAR-1995.
06-SEP-1994; 94WO-SE00826.
06-SEP-1993; 93SE-0002855.
(GUSS/) GUSS B.
(JONS/) JONSSON H.
(LIND/) LINDBERG M.
(MUEL/) MUELLER H.
(RANT/) RANTAMAKI L K.
Guss B, Jonsson H, Lindberg M, Mueller H, Rantamaki LK;
WPI: 1995-123382/16.
N-PSDB: AA089196.
DNA encoding fast alpha 2-macro:globulin-binding proteins - used
to obtain prods. for sepn., detection or quantification or for
binding inhibition
Disclosure: Page 31; 50pp; English.
A phage lambda GEM-11 library of S. dysgalactiae SCL DNA was analyzed
for fast-acting alpha-2-macroglobulin (FAM)- and IgG-binding
activity. One clone, which expressed both activities, was
analyzed to obtain DNA encoding the FAM-binding protein, MIG
(AAR71929). The mig gene is given in AA089196.
XX
SQ Sequence 664 AA:

Query Match 13.5%; Score 105; DB 16; Length 664;
Best Local Similarity 23.9%; Pred. No. 0.41;
Matches 54; Conservative 21; Mismatches 75; Indels 76; Gaps 9;

QY 1 VDSPIEQP-----RIIPNGGLTNLGNAPKELARNEERAIIDELKKOATED----- 47
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 423 IDAPELTPALTYTKLYKGNTFES---GETTTKAV--DAETAEKAFKOYANENGVGKWSY 477
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 48 KEATTAIEAASDALEALADOTDLOSEEAAYVKADNAASDALEALADOTDLOSEEA-E 106
   || : || : || : || : || : || : || : || : || : || : || : || : || :
DB 478 DDAKTFVTTEKRAVIDAPELTPALTYTKLYV-----NGKTLKGETTTRAVDAETAEKAFK 533
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 107 VVQSDNAASDAW-----EK-----AATPIADV 129
   : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 534 QYANENGVDGVWYDADTKTFVTTEMVTEVVGADAPTEPEKEASIPVPLTPATPIAKDD 593
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 130 KKTGDKPV-----VKKERONVNTLPTTGEESEN 159
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

```

DB 594 AKKDDTKKVDTKEDAKRPEAKKEAKKEAKKATLPTTGGSNP 639  
RESULT 40  
AACG37153  
ID AACG37153 standard; Protein: ISI AA.  
XX AAC37153;  
XX 18-OCT-2000 (first entry)  
XX Arabidopsis thaliana protein fragment SEQ ID NO: 45637.  
DE Protein identification; signal transduction pathway; metabolic pathway;  
KM hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX Arabidopsis thaliana.  
XX Ep1033405-A2.  
XX 06-SEP-2000.  
XX 25-FEB-2000; 2000EP-0301439.  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132407.  
PR 05-MAY-1999; 99US-0132484.  
PR 06-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132863.  
PR 14-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.

PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143624.  
PR 14-JUL-1999; 99US-0143542.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144086.  
PR 16-JUL-1999; 99US-0144088.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.

PR 18-AUG-1999; 9905-0149426.  
 PR 20-AUG-1999; 9905-0149722.  
 PR 20-AUG-1999; 9905-0149723.  
 PR 20-AUG-1999; 9905-0149929.  
 PR 23-AUG-1999; 9905-0149802.  
 PR 23-AUG-1999; 9905-0149930.  
 PR 25-AUG-1999; 9905-0150566.  
 PR 26-AUG-1999; 9905-0150884.  
 PR 27-AUG-1999; 9905-0151065.  
 PR 27-AUG-1999; 9905-0151066.  
 PR 27-AUG-1999; 9905-0151080.  
 PR 30-AUG-1999; 9905-0151303.  
 PR 31-AUG-1999; 9905-0151438.  
 PR 01-SEP-1999; 9905-0151930.  
 PR 07-SEP-1999; 9905-0152363.  
 PR 10-SEP-1999; 9905-0153070.  
 PR 13-SEP-1999; 9905-0153758.  
 PR 15-SEP-1999; 9905-0154018.  
 PR 16-SEP-1999; 9905-0154039.  
 PR 20-SEP-1999; 9905-0154779.  
 PR 22-SEP-1999; 9905-0155139.  
 PR 23-SEP-1999; 9905-0155486.  
 PR 24-SEP-1999; 9905-0155659.  
 PR 28-SEP-1999; 9905-0156458.  
 PR 29-SEP-1999; 9905-0156596.  
 PR 04-OCT-1999; 9905-0157117.  
 PR 05-OCT-1999; 9905-0157753.  
 PR 06-OCT-1999; 9905-0157865.  
 PR 07-OCT-1999; 9905-0158029.  
 PR 08-OCT-1999; 9905-0158232.  
 PR 12-OCT-1999; 9905-0158369.  
 PR 13-OCT-1999; 9905-0159293.  
 PR 13-OCT-1999; 9905-0159294.  
 PR 13-OCT-1999; 9905-0159295.  
 PR 14-OCT-1999; 9905-0159329.  
 PR 14-OCT-1999; 9905-0159330.  
 PR 14-OCT-1999; 9905-0159331.  
 PR 14-OCT-1999; 9905-0159637.  
 PR 14-OCT-1999; 9905-0159638.  
 PR 18-OCT-1999; 9905-0159584.  
 PR 21-OCT-1999; 9905-0160741.  
 PR 21-OCT-1999; 9905-0160767.  
 PR 21-OCT-1999; 9905-0160768.  
 PR 21-OCT-1999; 9905-0160770.  
 PR 21-OCT-1999; 9905-0160814.  
 PR 21-OCT-1999; 9905-0160815.  
 PR 22-OCT-1999; 9905-0160980.  
 PR 22-OCT-1999; 9905-0160981.  
 PR 22-OCT-1999; 9905-0160989.  
 PR 25-OCT-1999; 9905-0161404.  
 PR 25-OCT-1999; 9905-0161405.  
 PR 25-OCT-1999; 9905-0161406.  
 PR 26-OCT-1999; 9905-0161359.  
 PR 26-OCT-1999; 9905-0161360.  
 PR 26-OCT-1999; 9905-0161361.  
 PR 26-OCT-1999; 9905-0161920.  
 PR 28-OCT-1999; 9905-0161992.  
 PR 28-OCT-1999; 9905-0161993.  
 PR 29-OCT-1999; 9905-0162142.

Query Match 13.4%; Score 104.5; DB 21; Length 151;  
 Best Local Similarity 26.0%; Pred. No. 0.073;  
 Matches 39; Conservative 25; Mismatches 57; Indels 29; Gaps 6;

OY 12 PNCGLTINLGNAP-----EKLALNBERAIDELKQAIEDKEATTAIEAASDLEALA 66  
 Db 8 PKESDMQNEESGVNPKVESENVAKENNTESGEKONQTVATTETTSVEAKETFPVEPTK 67  
 OY 67 DQTDALQSE-----EAAVY---KADNAASDLEALADQTDALQSEFAEAYQVS 110  
 Db 68 ETTFAVQPEVAANAVERSSADAGEAAVAPEKVENAATENAENAEKAVEAANAVERVEY--- 124  
 OY 111 DNAAADAMEKA-ATPIALDVYKTK-DTKPV 138

Db 125 ---AVEAEKKAEAPVKAEEAPVKAEEAPV 151  
 Search completed: October 13, 2002, 04:45:33  
 Job time : 67.6203 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 13, 2002, 04:50:13 ; Search time 330.61 Seconds  
(without alignments)  
1143.085 Million cell updates/sec

Title: US-09-847-539a-6\_COPY\_59\_86  
Perfect score: 129  
Sequence: 1 SDALENDADQTDALDSEEAAYVKADNAA 28

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL-frame+ .p2n.model -DEV=slh  
-Q/cgn2\_1/USPTO.spool/US09847539/runat\_10102002\_093106\_5005/app\_query.fasta.1.526  
-DB=EST -QFMT=fastap -SUFFIX=std.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS-bits -START=1 -END=1 -MATRIX-blosum62 -TRANS-human40.cdi -LIST=45  
-DOCALLIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=40 -MODE=LOCAL  
-OUTFMT=pro -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09847539.ecgn\_1\_1\_763\_@runat\_10102002\_093106\_5005 -NCPU=6 -ICPU=3  
-NO\_XLPHY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:.\*  
1: em\_estdb:\*  
2: em\_esthum:\*  
3: em\_estlin:\*  
4: em\_estinu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_estl:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_inv:\*  
15: em\_gss\_pln:\*  
16: em\_gss\_vrl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	54	41.9	788	10	BM400696
2	54	41.9	861	10	BM396025

3	53	41.1	613	10	BM012116
4	53	41.1	615	10 <td>BM011078</td>	BM011078
5	53	41.1	624	10 <td>BM013747</td>	BM013747
6	53	41.1	1101	12	CNS00FX5
7	52	40.3	438	9	AA941632
8	52	40.3	553	9	BE056346
9	52	40.3	611	12	AG051191
10	52	40.3	1163	12	CNS06M17
11	51.5	39.9	682	10	BM411667
12	51	39.5	243	10	BF496893
13	51	39.5	333	9	AA140863
14	51	39.5	383	9	AI552971
15	51	39.5	387	9	AI465756
16	51	39.5	454	9	AA423358
17	51	39.5	491	9	AI405338
18	51	39.5	529	9	AI0101404
19	51	39.5	535	9	AA540164
20	51	39.5	548	9	AA659168
21	51	39.5	553	10	BM431282
22	51	39.5	558	9	AU206045
23	51	39.5	563	10	BF486089
24	51	39.5	567	9	AI063297
25	51	39.5	567	9	AI388871
26	51	39.5	567	10	BI637415
27	51	39.5	568	9	AI109571
28	51	39.5	569	9	AA392809
29	51	39.5	581	9	AI388559
30	51	39.5	583	9	AI108834
31	51	39.5	615	12	AZ698116
32	51	39.5	622	9	BB201301
33	51	39.5	633	9	AI062828
34	51	39.5	643	9	AA391497
35	51	39.5	650	10	BE337641
36	51	39.5	660	10	BI633216
37	51	39.5	666	9	AI062753
38	51	39.5	666	12	AZ699109
39	51	39.5	710	10	BI637204
40	51	39.5	797	10	BF494445
41	50.5	39.1	561	12	TA2618020
42	50	38.8	445	10	BG404048
43	50	38.8	483	10	BF487636
44	50	38.8	495	9	AU208069
45	50	38.8	528	9	AV393274

#### ALIGNMENTS

RESULT 1  
BM400696 788 bp mRNA linear EST 17-JAN-2002  
5009-0-77-F09.t.1 Chilcoat/Turkewitz CDNA (large fraction)  
Tetrahymena thermophila cDNA, mRNA sequence.

ACCESSION  
BM400696  
VERSION  
BM400696.1 GI:18200749  
KEYWORDS  
EST.

SOURCE  
ORGANISM  
Tetrahymena thermophila.

REFERENCE  
1. (bases 1 to 788)  
Turkewitz R.P., Karer, K.M., Jahn, C., Orlas, E., Kirk, K.E., Frankel  
J. and Klobutcher, L.

TITLE  
JOURNAL  
COMMENT  
EST from Tetrahymena thermophila, strain CU428.1, growing cells  
Unpublished (2002)  
Contact: Turkewitz AP  
Molecular Genetics and Cell Biology  
University of Chicago  
920 E. 58th Street, Chicago, IL 60637, USA  
Tel: 773 702 4374  
Fax: 773 702 3172  
Email: apturkew@midway.uchicago.edu  
Seq primer: T3.

FEATURES  
Location/Qualifiers

```

source
1. .788
/organism="Tetrahymena thermophila"
/strain="CU428.1"
/db_xref="taxon:5911"
/clone_lib="Chllicoat/Turkewitz cDNA (large fraction)"
/notes="Vector: Bluescript2 SK+, details on library
preparation can be found in Chllicoat and Turkewitz (2001)
Proc. Natl. Acad. Sci USA, 98: 8709-8713."
Proc. Natl. Acad. Sci USA, 98: 8709-8713."

BASE COUNT      248 a      219 c      121 g      199 t      1 others
ORIGIN

Alignment Scores:
Pred. NO.:      103      Length:      788
Score:          54.00      Matches:      12
Percent Similarity: 66.67% Conservative: 4
Best Local Similarity: 50.00% Mismatches: 8
Query Match:      41.86% Indels: 0
Db:              10      Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x BM400696 (1-788)

Oy      2      Aspa1aLeug1a1aLeua1aAsp1nThraspa1aLeug1aSerGlua1a1a1a1 21
         |||||      ::|||      ||      |||||      ::      ||      ||:::
Db      265      GATGCTCAACAGCCTCTAGACACAACACTGATGCTCAATGCTGCCGAGAACCTCTT 324
Oy      22      Val1ySa1aAsp 25
         |||||      ::|||      ||      |||||      ::      ||      ||:::
Db      325      GCCAATGCTGAA 336

RESULT 2
LOCUS      BM396025      861 bp      mRNA      linear      EST 17-JAN-2002
DEFINITION      5009-0-15-603.t.2 Chllicoat/Turkewitz cDNA (large fraction)
               Tetrahymena thermophila cDNA, mRNA sequence.
ACCESSION      BM396025
VERSION      BM396025.1      GI:18196078
KEYWORDS      EST.
SOURCE      Tetrahymena thermophila.
ORGANISM      Tetrahymena thermophila.
Eukaryota; Alveolata; Ciliophora; Oligohymenophorea;
Hymenostomolida; Tetrahymenida; Tetrahymena.
AUTHORS      Turkewitz,A.P., Karier,K.M., Jahn,C., Orlas,E., Kirk,K.E., Frankel
               ,J. and Klobutcher,L.
EST from Tetrahymena thermophila, strain CU428.1, growing cells
unpublished (2002)
CONTACT: Turkewitz AP
Molecular Genetics and Cell Biology
University of Chicago
920 E. 58th Street, Chicago, IL 60637, USA
Tel: 773 702 4374
Fax: 773 702 3172
Email: apurkew@midway.uchicago.edu
Seq primer: T3.

FEATURES
Source
1..861
location/Qualifiers
/organism="Tetrahymena thermophila"
/strain="CU428.1"
/db_xref="taxon:5911"
/clone_lib="Chllicoat/Turkewitz cDNA (large fraction)"
/notes="Vector: Bluescript2 SK+, details on library
preparation can be found in Chllicoat and Turkewitz (2001)
Proc. Natl. Acad. Sci USA, 98: 8709-8713."
Proc. Natl. Acad. Sci USA, 98: 8709-8713."

BASE COUNT      278 a      222 c      151 g      207 t      3 others
ORIGIN

Alignment Scores:
Pred. NO.:      117      Length:      861
Score:          54.00      Matches:      12
Percent Similarity: 66.67% Conservative: 4
Best Local Similarity: 50.00% Mismatches: 8
Query Match:      41.86% Indels: 0
Db:              10      Gaps: 0

```

```

US-09-847-539A-6_COPY_59_86 (1-28) x BW396025 (1-861)

Oy      2 Aspa1ala1eug1ua1aleua1aaspclntrrAspa1ala1euglnserc1ue1ua1a1aval 21
        ||||||| :|||:||||| ||| | ||||| ||||| ||| :||| |||:|
Db       501 GATGCTCAACAGCCTCTTCAAGACAACAACACTGATGCTTCATTAAATGCTCCGAGAAGACTCTT 560
        |||||:|

Oy      22 Val1ysAlaasp 25
        |||||:|
Db       561 GCCAATGCTGA 572

RESULT 3
LOCUS    Bf012116                               613 bp   mRNA          linear   EST 05-DEC-2001
DEFINITION Bf012116 MF01SSA cDNA Oryzias latipes cDNA clone MF01SSA171D09 5',
mRNA sequence.
ACCESSION Bf012116
VERSION    Bf012116.1 GI:17354928
KEYWORDS   EST.
SOURCE     Japanese medaka.
ORGANISM   Euryzias latipes
            Euteleostomi; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorphi; Acanthopterygii; Percormorpha; Atherinomorpha;
            Belontiiformes; Adrianiichthyidae; Cyprinidae; Oryziat.
REFERENCE 1 (bases 1 to 613)
AUTHORS   Kohara,Y., Shin-I,T., Kimura,T., Narita,T., Jinbo,T. and Takeda,H.
TITLE      Medaka EST Project In Takeda's Lab
JOURNAL    Unpublished (2001)
COMMENT    Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tsuhit@genes.nig.ac.jp.

FEATURES             source
                     1..613
                        /organism="Oryzias latipes"
                        /strain="Hg-rR"
                        /db_xref="taxon:8090"
                        /clone="MF01SSA171D09"
                        /clone_11b="MF01SSA cDNA"
                        /sex="mixture of female and male"
                        /tissue_type="whole embryo"
                        /dev_stage="segmentation stage 20 - 25"
BASE COUNT   134 a      136 c      239 g      104 t

ORIGIN
Alignment Scores:
Pred. No.:      104                      Length:      613
Score:           53.00                   Matches:      12
Percent Similarity: 60.71%               Conservative:  5
Best Local Similarity: 42.86%              Mismatches:   11
Query Match:     41.09%                  Indels:       0
DB:              10                       Gaps:         0

US-09-847-539A-6_COPY_59_86 (1-28) x Bf012116 (1-613)

Oy      1 Seraspa1ala1eug1ua1aleua1aaspclntrrAspa1ala1euglnserc1ue1ua1a1aval 20
        |||||:| ||| :|||:||||| ||| | ||||| ||||| ||| :||| |||:|
Db       8 TCAGAGCGCGCGCACGAGCGCGGCAGCAGCAGGAGGGCGCTTTAACCGAGAGAGCGCGCG 67
        |||||:|

Oy      21 ValValysAlaaspasna1a1aal 28
        |||||:|
Db       68 GCTTTAACCGAGAGAGCGCGCT 91

RESULT 4
LOCUS    Bf011078                               615 bp   mRNA          linear   EST 05-DEC-2001
DEFINITION Bf011078 MF01SSA cDNA Oryzias latipes cDNA clone MF01SSA153r11 5',
mRNA sequence.
ACCESSION Bf011078
```



```

VERSION      BU011078.1  GI:17367422
KEYWORDS     EST.
SOURCE       Japanese medaka.
ORGANISM     Oryzias latipes

REFERENCE    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
              Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
              Belontiiformes; Adrianichthyidae; Oryziinae; Oryzias.
AUTHORS      Kohara,Y., Shin-I,T., Kimura,T., Narita,T., Jindo,T. and Takeda,H.
TITLE        Medaka EST Project in Takeda's lab
JOURNAL      Unpublished (2001)
COMMENT      Contact: Tadasu Shin-I
              National Institute of Genetics
              1111 Yata, Mishima, Shizuoka 411-8540, Japan
              Tel: 81-559-81-6856
              Fax: 81-559-81-6855
              Email: tshin@genes.nig.ac.jp.

FEATURES
source       1..615
              /organism="Oryzias latipes"
              /strain="Hd-r"
              /db_xref="taxon:8090"
              /clone_1lb="MF01SSA153F11"
              /sex="mixture of female and male"
              /tissue_type="whole embryo"
              /dev_stage="segmentation stage 20 - 25"

BASE COUNT   135 a 137 c 234 g 109 t
ORIGIN
Alignment Scores:
Pred. No.:   105
Score:       53.00
Percent Similarity: 60.71%
Best Local Similarity: 42.86%
Query Match: 41.09%
DB:          10
              Length: 615
              Matches: 12
              Conservative: 5
              Mismatches: 11
              Indels: 0
              Gaps: 0

US-09-847-539a-6_COPY_59_86 (1-28) x BU011078 (1-615)
QY 1 SeraspalaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAla 20
    |||::|||::: |||||::: ||||| ||||| ||||| |||||
Db 40 TCAGAGCGCGCGGCGAGCGCGCGGAGAGGCGCTTTAACCGAGAGAGCGCGCG 99
QY 21 ValVallySAlaAspAsnAlaAla 28
    ::::: |||||
Db 100 GCTTTAACCGAGAGGAGCGGCT 123

RESULT 5
BU013747 624 bp mRNA linear EST 05-DEC-2001
LOCUS     BU013747 MF01SSA CDNA Oryzias latipes CDNA clone MF01SSA155A09 5',
          mRNA sequence.
ACCESSION BU013747
VERSION   BU013747.1 GI:17363622
KEYWORDS EST.
SOURCE    Japanese medaka.
ORGANISM  Oryzias latipes
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
          Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
          Belontiiformes; Adrianichthyidae; Oryziinae; Oryzias.
REFERENCE 1 (bases 1 to 624)
          Kohara,Y., Shin-I,T., Kimura,T., Narita,T., Jindo,T. and Takeda,H.
TITLE      Medaka EST Project in Takeda's lab
JOURNAL    Unpublished (2001)
COMMENT    Contact: Tadasu Shin-I
              Center For Genetic Resource Information
              National Institute of Genetics
              1111 Yata, Mishima, Shizuoka 411-8540, Japan
              Tel: 81-559-81-6856

```

```

Fax: 81-559-81-6855
Email: tshin@genes.nig.ac.jp.
Location/Qualifiers
1..624
/organism="Oryzias latipes"
/strain="Hd-r"
/db_xref="taxon:8090"
/clone_1lb="MF01SSA155A09"
/sex="mixture of female and male"
/tissue_type="whole embryo"
/dev_stage="segmentation stage 20 - 25"

BASE COUNT   144 a 139 c 232 g 108 t 1 others
ORIGIN
Alignment Scores:
Pred. No.:   107
Score:       53.00
Percent Similarity: 60.71%
Best Local Similarity: 42.86%
Query Match: 41.09%
DB:          10
              Length: 624
              Matches: 12
              Conservative: 5
              Mismatches: 11
              Indels: 0
              Gaps: 0

US-09-847-539a-6_COPY_59_86 (1-28) x BU013747 (1-624)
QY 1 SeraspalaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAla 20
    |||::|||::: |||||::: ||||| ||||| ||||| |||||
Db 65 TCAGAGCGCGCGGCGAGCGCGCGGAGAGGCGCTTTAACCGAGAGAGCGCGCG 124
QY 21 ValVallySAlaAspAsnAlaAla 28
    ::::: |||||
Db 125 GCTTTAACCGAGAGGAGCGGCT 148

RESULT 6
CNS00EFX5/c 1101 bp DNA linear GSS 03-JUN-1999
LOCUS      Drosophila melanogaster genome survey sequence T7 end of BAC:
DEFINITION BACR32D16 of RPCI-98 library from Drosophila melanogaster (fruit
fly) genomic survey sequence.
ACCESSION  AL071361
VERSION     AL071361.1 GI:4951201
KEYWORDS   GSS.
SOURCE      fruit fly.
ORGANISM   Drosophila melanogaster
          Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
          Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
          Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE  1 (bases 1 to 1101)
          Genoscope.
          Direct Submission
          Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage;
          BP 191 91006 Evry cedex - FRANCE (E-mail: seqref@genoscope.cns.fr
          - Web : www.genoscope.cns.fr)
          Determination of this BAC-end sequence was carried out as part of a
          collaboration with the Berkeley Drosophila Genome Project (BDGP).
          The BDGP is constructing a physical map of the Drosophila
          melanogaster genome using these BACs. For further information
          please see http://www.fruitfly.org The BDGP Osoegawa and
          melanogaster BAC library was prepared by Kazutoyo Osoegawa and
          Aaron Mammoseer in Pieter de Jong's laboratory in the Department of
          Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
          NY. The library is named RPCI-98 and was constructed by partial
          EcoRI digestion of Drosophila DNA provided by the BDGP from the
          isogenic strain y2; cn bw sp; the same strain used for the BDGP's
          P1 and EST libraries. A more detailed description of the library
          and how to order individual BAC clones, the entire library, or
          filters for hybridization from the BACPAC Resource Center can be
          found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES
source       1..1101
              /organism="Drosophila melanogaster"
              /db_xref="taxon:7227"
              /clone_1lb="RPCI-98"

```

BASE COUNT 268 a 329 c 226 g 231 t 47 others  
 ORIGIN

## Alignment Scores:

Pred. No.: 236 Length: 1101  
 Score: 53.00 Matches: 12  
 Percent Similarity: 66.67% Conservative: 4  
 Best Local Similarity: 50.00% Mismatches: 8  
 Query Match: 41.09% Indels: 0  
 DB: 12 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x CNS00FX5 (1-1101)

QY 5 GIUAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValAlaLysAla 24  
 DB 263 GAGGAGAGAGCTGACGAGGAGAGCCGCCGAGGAGAGATGCCGANNAGCTGCCGCC 204

QY 25 ASPAaAlaAla 28  
 DB 203 GATGCTGCCGCC 192

## RESULT 7

AA941632 438 bp mRNA linear EST 19-APR-2001  
 LOCUS LD25873.Sprime LD Drosophila melanogaster embryo POT2 Drosophila  
 DEFINITION melanogaster cDNA clone LD25873 Sprime, mRNA sequence.

ACCESSION AA941632  
 VERSION AA941632.1 GI:3101545  
 KEYWORDS EST.  
 SOURCE fruit fly.  
 ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 438)  
 Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,  
 Lewis,S. and Rubin,G.M.  
 BDGP/HMI Drosophila EST Project  
 TITLE Unpublished (2001)  
 JOURNAL Contact: Stapleton, M.  
 COMMENT BDGP

Lawrence Berkeley National Lab  
 One Cyclotron Rd, Berkeley, CA 94720, USA  
 Fax: 510 486 6798

Email: <http://www.fruitfly.org/EST/estfruitfly.berkeley.edu>  
 Plate: 258 row: G column: 1  
 High quality sequence stop: 226.

## FEATURES

source 1..438  
 Location/Qualifiers  
 /organism="Drosophila melanogaster"  
 /db\_xref="taxon:7227"  
 /clone="LD25873"  
 /clone\_11b="LD Drosophila melanogaster embryo POT2"  
 /sex="male and female"  
 /dev\_stage="0 to 24 hours mixed stage embryonic"  
 /lab\_host="X11 Blue"  
 /note="Organ: embryo. Vector: POT2. Site\_1: EcoRI; Site\_2:  
 XhoI; Sized fractionated cDNAs were directly ligated into  
 POT2."

BASE COUNT 98 a 125 c 99 g 116 t  
 ORIGIN

## Alignment Scores:

Pred. No.: 93.8 Length: 438  
 Score: 52.00 Matches: 11  
 Percent Similarity: 55.56% Conservative: 4  
 Best Local Similarity: 40.74% Mismatches: 12  
 Query Match: 40.31% Indels: 0  
 DB: 9 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x AA941632 (1-438)

QY 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAla 20  
 DB 231 AACGATCGCGGCAATCATTCATAGACCCACAGACGAGAGACGAGACAGACAGCG 172

QY 21 ValValLysAlaAspAsnAla 27  
 DB 171 CGCGTACGAGAGATCGTCC 151

## RESULT 8

BE056346 553 bp mRNA linear EST 08-JUN-2000  
 LOCUS 00257 leafy spurge Lambda HybridZAP 2.1 two-hybrid vector cDNA  
 DEFINITION library Euphorbia esula cDNA clone 25D 5' similar to Arabidopsis  
 unknown Protein (accession# AAF24521), mRNA sequence.

ACCESSION BE056346  
 VERSION BE056346.1 GI:8383403

KEYWORDS EST.  
 SOURCE leafy spurge.  
 ORGANISM Euphorbia esula

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Rosidae; eurosids I; Malpighiales; Euphorbiaceae; Euphorbia.

REFERENCE 1 (bases 1 to 553)  
 Anderson,J.V. and Horvath,D.P.  
 Identification of mRNAs expressed in underground adventitious buds  
 of Euphorbia esula (leafy spurge)  
 TITLE Unpublished (2000)  
 JOURNAL Contact: Anderson JV  
 COMMENT Plants Science Research  
 USDA/ARS, Biosciences Research Lab  
 1605 Albrecht Blvd., PO Box 5674, Fargo, ND 58105, USA  
 Tel: 701 239 1263  
 Fax: 701 239 1252  
 Email: [andersjv@fargo.ars.usda.gov](mailto:andersjv@fargo.ars.usda.gov)  
 Seq primer: PADS.

## FEATURES

source 1..553  
 Location/Qualifiers  
 /organism="Euphorbia esula"  
 /db\_xref="taxon:3993"  
 /clone="25D"  
 /clone\_11b="leafy spurge Lambda HybridZAP 2.1 two-hybrid  
 vector cDNA library"  
 /tissue-type="underground adventitious buds"  
 /dev\_stage="3-day induced (decapitated)"

BASE COUNT 161 a 117 c 116 g 159 t  
 ORIGIN

## Alignment Scores:

Pred. No.: 130 Length: 553  
 Score: 52.00 Matches: 12  
 Percent Similarity: 65.38% Conservative: 5  
 Best Local Similarity: 46.15% Mismatches: 9  
 Query Match: 40.31% Indels: 0  
 DB: 9 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x BE056346 (1-553)

QY 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAla 20  
 DB 541 TCAGAGCACCTTCTTCACTGCTGATGCCACAAAGCGATGATCAGACCACTTGCC 482

QY 21 ValValLysAlaAspAsn 26  
 DB 481 ATTTCATACACGATAT 464

## RESULT 9

AO951191 611 bp DNA linear GSS 27-JAN-2000  
 LOCUS Sheared DNA-50A17.TR Sheared DNA trypanosoma brucei genomic clone  
 DEFINITION Sheared DNA-50A17, DNA sequence.  
 ACCESSION AO951191  
 VERSION AO951191.1 GI:6774456



VERSION BM441667.1 GI:18472442  
KEYWORDS EST.  
SOURCE Barley.  
ORGANISM Hordeum vulgare  
Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae  
; Triticeae; Hordeum.  
1 (bases 1 to 682)  
REFERENCE Hedley, P., Liu, H., Caldwell, D., McCallum, N., Mudie, S., Cardle, L., Ramsay, L., Machray, G., Marshall, D.F.M. and Waugh, R.  
TITLE Development of Barley Transcriptome Resources  
JOURNAL Unpublished (2001)  
COMMENT Contact: Waugh R  
Unit of Genomics  
Scottish Crop Research Institute  
Invergowrie, Dundee, DD2 5DA, Scotland, UK  
Tel: 00 44 1382 562731  
Fax: 00 44 1382 562426  
Email: rwaugh@scri.sari.ac.uk  
All sequence has a Phred quality score of 20 or over  
Seq primer: M13 reverse.

FEATURES  
source  
1..682  
/organism="Hordeum vulgare"  
/cultivar="Optic"  
/db\_xref="taxon:4513"  
/clone="EBed07\_S0001.D10"  
/clone\_lib="IGF Barley EBed07 library"  
/tissue\_type="Endosperm"  
/dev\_stage="28 days post anthesis"  
/lab\_host="DH10B"  
/note="Vector: pSPORT1; Site\_1: Sal I; Site\_2: Not I;  
Non-normalised library, directionally cloned into pSPORT1.  
Derived from endosperm tissue dissected from developing  
grains (28 days post anthesis) in glasshouse grown barley  
plants. Developed as part of the barley transcriptome  
resources of BBSRC/SEERAD funded cereal IGF (Investigating  
Gene Function) project."

BASE COUNT 117 a 241 c 211 g 113 t  
ORIGIN

Alignment Scores:  
Pred. No.: 208 Length: 682  
Score: 51.50 Matches: 17  
Percent Similarity: 67.86% Conservative: 2  
Best Local Similarity: 60.71% Mismatches: 8  
Query Match: 39.92% Indels: 1  
DB: 10 Gaps: 1

US-09-847-539a-6\_COPY\_59\_86 (1-28) x BM441667 (1-682)

QY 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluAlaAla 20  
||| ::| ||||| ||| ||||| ||||| ::| |||||  
DB 427 TCCTGCTGCGAGAGAGCCTTGCCCGCATGACAGACGCCCTG--AGCAGATGCCGCG 371

QY 21 ValValLysAlaAspAsnAlaAla 28  
||||| ||| |||||  
DB 370 GTCGCTCTCTGATGATGTCGACGCG 347

RESULT 12  
LOCUS BF496493/c 243 bp mRNA linear EST 19-APR-2001  
DEFINITION At10425.Sprime AT Drosophila melanogaster adult testes POTB7.  
Drosophila melanogaster cDNA clone AT10425 5 similar to CG7289;  
Fban0007289 located on: 2L 22B8-22B8:: 04/08/2001, mRNA sequence.  
ACCESSION BF496493 GI:13694046  
VERSION BF496493.2  
KEYWORDS EST.  
SOURCE fruit fly.  
ORGANISM Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 243)  
AUTHORS Stapleton, M., Brockstein, P., Hong, L., Agdayan, A., Baxter, E., Berman  
B., Carlson, J., Champe, M., Chavez, C., Chew, M., Dorsett, V., Farfan  
D., Frise, E., George, R., Gonzalez, M., Guatin, H., Harris, N., Li, P.,  
Liao, G., Miranda, A., Misra, S., Mungall, C.J., Nuno, J., Pacib, J.,  
Paragas, V., Park, S., Phouanavong, S., Wan, K., Yu, C., Lewis, S.E.,  
Celniker, S. and Rubin, G.M.  
TITLE BDGP/HMI AT Drosophila EST Project  
JOURNAL Unpublished (2000)  
COMMENT On Dec 6, 2000 this sequence version replaced gi:11579867.  
Contact: Stapleton, M.  
BDGP  
Lawrence Berkeley National Lab  
One Cyclotron Rd, Berkeley, CA 94720, USA  
Fax: 510 486 6798  
Email: [http://www.fruitfly.org/EST\\_estefruitfly.berkeley.edu](http://www.fruitfly.org/EST_estefruitfly.berkeley.edu)  
Plate: AT104 row: C column: 1  
High quality sequence stop: 235.

FEATURES  
source  
1..243  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/clone="AT10425"  
/clone\_lib="AT Drosophila melanogaster adult testes POTB7"  
/sex="male"  
/dev\_stage="0-3 day old Ore-R males"  
/lab\_host="Plates AT\_10-AT\_120: DHS-alpha. Plates  
AT\_121-AT\_319: DHS-alpha TONA"  
/note="Organ: ADULT testes; Vector: POTB7; Site\_1: EcoRI;  
Site\_2: XhoI; The mRNA for the testis library was made  
from testes and seminal vesicles hand dissected from 0-3  
day old Ore-R males. RNA kindly provided by the lab of  
Margaret Fuller. Sized fractionated cDNAs were directly  
ligated into POTB7. Plasmid cDNA library."

BASE COUNT 53 a 65 c 56 g 68 t  
ORIGIN

Alignment Scores:  
Pred. No.: 59.3 Length: 243  
Score: 51.00 Matches: 11  
Percent Similarity: 53.85% Conservative: 3  
Best Local Similarity: 42.31% Mismatches: 12  
Query Match: 39.53% Indels: 0  
DB: 10 Gaps: 0

US-09-847-539a-6\_COPY\_59\_86 (1-28) x BF496493 (1-243)

QY 2 AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluAlaAlaVal 21  
||| ::| ||| ||||| ||| ||||| ||||| ::| |||||  
DB 228 GATCGCGGAGATCATATGAGACACACAGCAGAGAGACAGAGACAGACAGCGCG 169

QY 22 ValLysAlaAspAsnAla 27  
||||| ||| |||||  
DB 168 GTACGACGAGATCGTCC 151

RESULT 13  
LOCUS AA140863/c 333 bp mRNA linear EST 29-NOV-1998  
DEFINITION CK00525.Sprime CK Drosophila melanogaster embryo blueScript.  
Drosophila melanogaster cDNA clone CK00525 Spprime, mRNA sequence.  
ACCESSION AA140863 GI:1704336  
VERSION AA140863.1  
KEYWORDS EST.  
SOURCE fruit fly.  
ORGANISM Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
REFERENCE Kopczyński, C., Serano, T., Rubin, G. and Goodman, C.  
AUTHORS BDGP/HMI CK Drosophila EST Project  
TITLE Unpublished (1996)  
COMMENT Other\_ESTs: CK00525.Sprime



FEATURES	Location/Qualifiers
source	1. .454

**FEATURES**  
**source**

```
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="GH25307"
```

```

/clone.lib="GH Drosophila melanogaster head POT2"
/sex="male and female"
/dev.stage="adult"
/lab.host="DH5 - alpha"
/notes="Organ: head; Vector: POT2; Site_1: EcoRI; Site_2:
XhoI; Sized fractionated cDNAs were directly ligated into
POT2. Plasmid cDNA library."
BASE COUNT      110 a      143 c      116 g      122 t
ORIGIN

Alignment Scores:
Pred. No.:      158      Length:      491
Score:          51.00      Matches:      11
Percent Similarity: 53.85%      Conservative: 3
Best Local Similarity: 42.31%      Mismatches: 12
Query Match:      39.53%      Indels:      0
DB:                9      Gaps:      0

US-09-847-539A-6_COPY_59_86 (1-28) x AA405338 (1-491)
QY      2  AspaLaLeuglUaLaLeuaLaaspgLnThAsPaLaLeugLnserGluGluAlaLaVal 21
      |||      |||      |||      |||      |||      |||
Db      227 GATCGGCGGGAATCATATAGAACACACAGACGAGGAGCAGAGGAACAACGCGCG 168
QY      22  VallysAlaAspaSnAla 27
      |||:::  |||  |||
Db      167 GTACGAGAGATCGTCGCC 150

RESULT 18
AUI01404      529 bp      mRNA      linear      EST 22-AUG-2000
LOCUS      AUI01404 Rice panicle shorter than 3cm Oryza sativa cDNA clone
DEFINITION      E30571, mRNA sequence.
ACCESSION      AUI01404
VERSION      AUI01404.1 GI:9865654
KEYWORDS      EST.
SOURCE      Oryza sativa.
ORGANISM      Oryza sativa.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE      1 (bases 1 to 529)
AUTHORS      Sasaki,T. and Yamamoto,K.
TITLE      Rice cDNA from panicle (2000)
JOURNAL      Unpublished (2000)
COMMENT      Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki
305-8602, Japan
Tel: 81-298-38-7441
Fax: 81-298-38-7468
Email: tsasaki@abr.affrc.go.jp, URL:http://rpg.dna.affrc.go.jp/
PROJECT="RGP".
E30571_7A.

FEATURES
      source
      1..529
      /organism="Oryza sativa"
      /strain="Nipponbare"
      /db_xref="taxon:4530"
      /clone="E30571"
      /clone.lib="Rice panicle shorter than 3cm"
      /dev.stage="shorter than 3cm"
      /note="Organ: panicle"
BASE COUNT      79 a      163 c      198 g      89 t
ORIGIN

Alignment Scores:
Pred. No.:      175      Length:      529
Score:          51.00      Matches:      13
Percent Similarity: 57.69%      Conservative: 2
Best Local Similarity: 50.00%      Mismatches: 11
Query Match:      39.53%      Indels:      0
DB:                9      Gaps:      0

```

```

US-09-847-539A-6_COPY_59_86 (1-28) x AUI01404 (1-529)
QY      3  AlLeuglUaLaLeuaLaaspgLnThAsPaLaLeugLnserGluGluAlaLaVal 22
      |||      |||      |||      |||      |||      |||
Db      216 GCGCGGAGCGCGGCTGACGAGAGGAGCGCGGCTACGAGAGGAGCGGAGCAGCG 275
QY      23  LysAlaAspaSnAlaLa 28
      |||||  |||||
Db      276 GCGCGGATACGCGCGCG 293

RESULT 19
AAS40164/c      535 bp      mRNA      linear      EST 19-APR-2001
LOCUS      AAS40164/c
DEFINITION      LD19555.5prime LD Drosophila melanogaster embryo Bluescript
Drosophila melanogaster cDNA clone LD19555 5prime, mRNA sequence.
ACCESSION      AAS40164
VERSION      AAS40164.1 GI:2286598
KEYWORDS      EST.
SOURCE      fruit fly.
ORGANISM      Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 535)
REFERENCE      Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
AUTHORS      Lewis,S. and Rubin,G.M.
TITLE      BDGP/HMI Drosophila EST Project
JOURNAL      Unpublished (2001)
COMMENT      Contact: Stapleton, M.
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
Plate: 195 row: E column: 7
High quality sequence stop: 411.

FEATURES
      source
      1..535
      /organism="Drosophila melanogaster"
      /db_xref="BDGP_EST:BDCl018642"
      /db_xref="taxon:7227"
      /clone="LD19555"
      /clone.lib="LD Drosophila melanogaster embryo Bluescript"
      /sex="male and female"
      /dev.stage="0 to 24 hours mixed stage embryonic"
      /lab.host="SOLR"
      /note="Organ: embryo; Vector: Bluescript SK; Site_1: EcoRI
      ; Site_2: XhoI; Constructed using Stratagene ZAP-cDNA
      Synthesis kit. Oligo dt-primed and directionally cloned at
      EcoRI and XhoI in Bluescript SK(+/-)"
BASE COUNT      117 a      153 c      131 g      134 t
ORIGIN

Alignment Scores:
Pred. No.:      178      Length:      535
Score:          51.00      Matches:      11
Percent Similarity: 53.85%      Conservative: 3
Best Local Similarity: 42.31%      Mismatches: 12
Query Match:      39.53%      Indels:      0
DB:                9      Gaps:      0

US-09-847-539A-6_COPY_59_86 (1-28) x AAS40164 (1-535)
QY      2  AspaLaLeuglUaLaLeuaLaaspgLnThAsPaLaLeugLnserGluGluAlaLaVal 21
      |||      |||      |||      |||      |||      |||
Db      211 GATCGGCGGGAATCATATAGAACACACAGACGAGGAGCAGAGGAACAACGCGCG 152
QY      22  VallysAlaAspaSnAla 27
      |||:::  |||  |||
Db      151 GTACGAGAGATCGTCGCC 134

RESULT 20

```

AM659168/c 548 bp mRNA linear EST 25-APR-2001  
LOCUS AM659168  
DEFINITION 96120 MARC 1BOV Bos taurus cDNA 5', mRNA sequence.  
ACCESSION AM659168  
VERSION AM659168.1 GI:7424995  
KEYWORDS EST.  
SOURCE  
ORGANISM Bos taurus  
COW.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
Bovidae; Bovinae; Bos.  
REFERENCE  
AUTHORS Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,  
Cassas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett,  
G.L., Heaton,M.P., Laegreid,W.M., Rohrer,G.A., Chitko-McKown,C.G.,  
Petee,G., Holt,L.I., Karameycheva,S., Liang,F., Quackenbush,J. and  
Keefe,J.W.  
TITLE Sequence evaluation of four pooled-tissue normalized bovine cDNA  
libraries and construction of a gene index for cattle  
JOURNAL Genome Res. 11 (4), 626-630 (2001)  
MEDLINE 21180013  
COMMENT Contact: Smith TPL  
USDA, ARS, US Meat Animal Research Center  
PO Box 166, Clay Center, NE 68933-0166, USA  
Tel: 402 762 4366  
Fax: 402 762 4390  
Email: smithemall.marc.usda.gov  
Single pass sequencing. Bases called and alt-trimmed with phred  
v0.980904.e. Vector identified by cross-match with the -minscore 18  
and -mismatch 12 options.  
PCR Primers  
FORWARD: AGGAACAGCTATGACCAT  
BACKWARD: GTTTCCTCAGTACAGC  
Plate: 83 row: M column: 13  
Seq primer: ATTTAGTGACATATAG.  
FEATURES  
SOURCE Location/Qualifiers  
1..548  
/organism="Bos taurus"  
/db\_xref="taxon:9913"  
/clone\_lib="MARC 1BOV"  
/tissue\_type="pooled"  
/lab\_host="DH10B"  
/note="Vector: pCMV SPORT6; Site\_1: XbaI; Site\_2: XhoI;  
Library made from pooled tissue from lymph node, ovary,  
fat, hypothalamus, and pituitary."  
BASE COUNT 132 a 143 c 151 g 132 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 184 Length: 548  
Score: 51.00 Matches: 11  
Percent Similarity: 60.87% Conservative: 3  
Best Local Similarity: 47.83% Mismatches: 9  
Query Match: 39.53% Indels: 0  
DB: 9 Gaps: 0  
US-09-847-539A-6\_COPY\_59\_86 (1-28) x AM659168 (1-548)  
QY 4 leuGuaLaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValys 23  
DB 538 CTTTCTCTCTTGTGCTCCAGGCTGATGCTCCAGGAAGAGAGAGACCCACCTGCG 479  
QY 24 AlaAspAsn 26  
DB 478 GAGGATGAC 470  
RESULT 21  
BM431282 553 bp mRNA linear EST 31-JAN-2002  
LOCUS BM431282  
DEFINITION J01012E08 Bos taurus Duodenum #1 library Bos taurus cDNA, mRNA  
sequence.  
ACCESSION BM431282  
VERSION BM431282.1 GI:18453004

KEYWORDS EST.  
SOURCE COW.  
ORGANISM Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
Bovidae; Bovinae; Bos.  
REFERENCE  
AUTHORS Hansen,C., Fu,A., Meng,Y., Li,C., Okine,E., Sensen,C.W., Gordon  
P.M.K. and Moore,S.S.  
TITLE Gene expression profiling of the Bovine Gastrointestinal Tract  
JOURNAL Unpublished (2002)  
COMMENT Contact: Dr. Stephen Moore  
Beef Genomics Laboratory  
Dept of AFNS, University of Alberta  
410 Agri/For, Dept of AFNS, U of A, Edmonton, AB, T6G 2P5, Canada  
Tel: 780 492 0169  
Fax: 780 492 4265  
Email: smoores@atns.ualberta.ca  
Insert Length: 553 Std Error: 0.00  
POLYA-No.  
FEATURES  
SOURCE Location/Qualifiers  
1..553  
/organism="Bos taurus"  
/db\_xref="taxon:9913"  
/clone\_lib="Bos taurus Duodenum #1 library"  
/tissue\_type="Smooth muscle"  
/cell\_type="Simple columnar epithelial"  
/dev\_stage="Young adult"  
/lab\_host="XLI-BlueMRF'strain"  
/note="Organ: Intestine/Duodenum; Vector: Uni-22APYR;  
Site\_1: EcoRI; Site\_2: Xho I"  
BASE COUNT 158 a 124 c 139 g 132 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 186 Length: 553  
Score: 51.00 Matches: 11  
Percent Similarity: 60.87% Conservative: 3  
Best Local Similarity: 47.83% Mismatches: 9  
Query Match: 39.53% Indels: 0  
DB: 10 Gaps: 0  
US-09-847-539A-6\_COPY\_59\_86 (1-28) x BM431282 (1-553)  
QY 4 leuGuaLaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValys 23  
DB 142 CTTTCTCTCTTGTGCTCCAGGCTGATGCTCCAGGAAGAGAGAGACCCACCTGCG 83  
QY 24 AlaAspAsn 26  
DB 82 GAGGATGAC 74  
RESULT 22  
AU206045 558 bp mRNA linear EST 17-JUL-2001  
LOCUS AU206045  
DEFINITION AU206045 unpublished oligo-capped cDNA library, stage L1  
Caenorhabditis elegans cDNA clone yk58e11 5', mRNA sequence.  
ACCESSION AU206045  
VERSION AU206045.1 GI:14838902  
KEYWORDS EST.  
SOURCE Caenorhabditis elegans.  
ORGANISM Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilita; Rhabditoides  
; Rhabditidae; Poloderinae; Caenorhabditis.  
REFERENCE  
AUTHORS Kohara,Y., Shin-I,T., Thierry-Mieg,J., Thierry-Mieg,D., Suzuki,Y.  
and Sugano,S.  
TITLE A complementary view of the C.elegans genome  
JOURNAL Unpublished (2001)  
COMMENT Contact: Yuji Kohara  
Genome Biology Lab.  
National Institute of Genetics  
Yata 1111, Mishima, Shizuoka 411, Japan



Tel: 81-559-81-6854

Fax: 81-559-81-6855

Email: ykohara@lab.nig.ac.jp.

## FEATURES

Location/Qualifiers

1. .558

/organism="Caenorhabditis elegans"

/strain="N2"

/db\_xref="taxon:6239"

/clone="yK858ell"

/clone\_lib="unpublished oligo-capped cDNA library, stage 1"

/sex="Hermaphrodite"

/tissue.type="whole animal"

/dev\_stage="L1"

BASE COUNT 164 a 146 c 142 g 105 t 1 others

ORIGIN

## Alignment Scores:

Pred. No.:	189	Length:	558
Score:	51.00	Matches:	9
Percent Similarity:	68.00%	Conservative:	8
Best Local Similarity:	36.00%	Mismatches:	8
Query Match:	39.53%	Indels:	0
DB:	9	Gaps:	0

US-09-847-539a-6\_COPY\_59\_86 (1-28) x AU206045 (1-558)

OY 1 SeraspalaleuglualaleualaaspclnThraspalaleuglnsercluglualala 20

Db 236 TCACATGCTATTCAAGACCTTCGATCAATTCAGCAACACTCAAAAGCAAAAGGAGAGA 295

OY 21 ValVallysAlaasp 25

Db 296 ATTGAGAGAGAGAA 310

RESULT 23

BF486089/c

LOCUS 563 bp mRNA linear EST 19-APR-2001

DEFINITION AT20267.5prime AT Drosophila melanogaster adult testes POTB7;

Drosophila melanogaster cDNA clone AT20267 5 similar to CG7289;

Phan0007289 located on: 2L 22B8-22B8:: 04/09/2001, mRNA sequence.

ACCESSION BF486089

VERSION BF486089.2 GI:13697978

KEYWORDS EST.

SOURCE fruit fly.

ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 563)

REFERENCE Stapleton, M., Brokstein, P., Hong, L., Abdayani, A., Baxter, E., Berman

, B., Carlson, J., Champe, M., Chavez, C., Chew, M., Dorsett, V., Farfan

, D., Frise, E., George, R., Gonzalez, M., Guarin, H., Harrel, N., Li, P.,

Liao, G., Miranda, A., Misra, S., Mungall, C.J., Nuno, J., Pacle, J.,

Paragas, V., Park, S., Phouenavong, S., Wan, K., Yu, C., Lewis, S.E.,

Celinker, S., and Rubin, G.M.

BDGP/HMT AT Drosophila EST Project

Unpublished (2000)

On Dec 6, 2000 this sequence version replaced gi:11569390.

COMMENT Contact: Stapleton, M.

BDGP

TITLE Lawrence Berkeley National Lab

One Cyclotron Rd, Berkeley, CA 94720, USA

Fax: 510 486 6798

Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu

hit genomic AE003584: arm:2L [1824960,2149443]

estimated-cyto:22B4-22D2: 04/09/2001

Plate: AT 202 row: F column: 7

High quality sequence stop: 562.

Location/Qualifiers

1. .563

/organism="Drosophila melanogaster"

/db\_xref="taxon:7227"

BASE COUNT 127 a 161 c 141 g 138 t

/clone="AT20267"

/clone\_lib="AT Drosophila melanogaster adult testes POTB7"

/sex="male"

/dev\_stage="0-3 day old Ore-R males"

/lab\_host="Plates AT-10-AT-120: DH5-alpha. Plates

AT-121-AT-319: DH5-alpha Tona"

/note="Organ: ADULT testes; Vector: POTB7; Site:1: EcoRI;

Site-2: XhoI. The mRNA for the testis library was made

from testes and seminal vesicles hand dissected from 0-3

day old Ore-R males. RNA kindly provided by the lab of

Margaret Fuller. Sized fractionated cDNAs were directly

ligated into POTB7. Plasmid cDNA library."

BASE COUNT 126 a 159 c 141 g 137 t

ORIGIN

Alignment Scores:

Pred. No.:	191	Length:	563
Score:	51.00	Matches:	11
Percent Similarity:	53.85%	Conservative:	3
Best Local Similarity:	42.31%	Mismatches:	12
Query Match:	39.53%	Indels:	0
DB:	10	Gaps:	0

US-09-847-539a-6\_COPY\_59\_86 (1-28) x BF486089 (1-563)

OY 2 AspalaleuglualaleualaaspclnThraspalaleuglnsercluglualalaval 21

Db 228 GATCGCGCGAATCAATATGAGACACACAGCAGAGACGAGAGAGACAGACAGCGCG 169

OY 22 ValVysAlaaspAAla 27

Db 168 GTACGAGAGATCGTCC 151

RESULT 24

A1063297/c

LOCUS 567 bp mRNA linear EST 19-APR-2001

DEFINITION GH03030.5prime GH Drosophila melanogaster head POT2 Drosophila

melanogaster cDNA clone GH03030 5prime, mRNA sequence.

ACCESSION A1063297

VERSION A1063297.1 GI:3339241

KEYWORDS EST.

SOURCE fruit fly.

ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 567)

REFERENCE Harvey, D., Brokstein, P., Hong, L., Evans-Holm, M., Su, C., Tsang, G.,

Lewis, S., and Rubin, G.M.

BDGP/HMT Drosophila EST Project

Unpublished (2001)

Contact: Stapleton, M.

BDGP

TITLE Lawrence Berkeley National Lab

One Cyclotron Rd, Berkeley, CA 94720, USA

Fax: 510 486 6798

Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu

hit genomic sequence DS06378

Plate: 30 row: C column: 6

High quality sequence stop: 454.

Location/Qualifiers

1. .567

/organism="Drosophila melanogaster"

/db\_xref="taxon:7227"

/clone="GH03030"

/clone\_lib="GH Drosophila melanogaster head POT2"

/sex="male and female"

/dev\_stage="adult"

/lab\_host="DH5-alpha"

/note="Organ: head; Vector: POT2; Site:1: EcoRI; Site-2:

XhoI. Sized fractionated cDNAs were directly ligated into

POT2. Plasmid cDNA library."

```

ORIGIN
Alignment Scores:
Pred. No.: 193
Score: 51.00
Percent Similarity: 53.85%
Best Local Similarity: 42.31%
Query Match: 39.53%
DB: 9

US-09-847-539a-6_COPY_59_86 (1-28) x A1063297 (1-567)
OY 2 AspalaleuGlua1aleuAlaaspGIntThraspalaleuGInserGIua1ala1aVal 21
Db 228 GATCGCGCGAATCATATAGAACACACAGACGAGAGAGAGGAAACAAGCGCG 169
OY 22 VallysAlaaspasna1a 27
Db 168 GTACGAGAGATCTGCTGCC 151

RESULT 25
A1388871/c 567 bp mRNA linear EST 19-APR-2001
LOCUS GH19913.5prtime GH Drosophila melanogaster head pot2 Drosophila
DEFINITION melanogaster cDNA clone GH19913 5prtime, mRNA sequence.
ACCESSION A1388871
VERSION A1388871.1 GI:4202882
KEYWORDS EST.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 567)
Harvey, D., Brokstein, P., Hong, L., Evans-Holm, M., Su, C., Tsang, G.,
Lewis, S. and Rubin, G. M.
BDGP/HM1 Drosophila EST project
Unpublished (2001)
Contact: Stapleton, M.

TITLE Lawrence Berkeley National Lab
JOURNAL One Cyclotron Rd, Berkeley, CA 94720, USA
COMMENT Fax: 510 486 6798
Email: http://www.fruitfly.org/EST, east@fruitfly.berkeley.edu
hit genomic sequence AC005749
Plate: 199 row: B column: 1
High quality sequence stop: 566.
Location/Qualifiers
1..567
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="GH Drosophila melanogaster head pot2"
/sex="male and female"
/dev_stage="adult"
/lab_host="DHS - alpha"
/notes="Organ: head; Vector: pot2; Site_1: EcoRI; Site_2:
XhoI; Sized fractionated cDNAs were directly ligated into
pot2. Plasmid cDNA library."

BASE COUNT 127 a 161 c 141 g 138 t
ORIGIN
Alignment Scores:
Pred. No.: 193
Score: 51.00
Percent Similarity: 53.85%
Best Local Similarity: 42.31%
Query Match: 39.53%
DB: 9

US-09-847-539a-6_COPY_59_86 (1-28) x A1388871 (1-567)
OY 2 AspalaleuGlua1aleuAlaaspGIntThraspalaleuGInserGIua1ala1aVal 21

```

```

Db 227 GATCGCGCGAATCATATAGAACACACAGACGAGAGAGAGGAAACAAGCGCG 168
OY 22 VallysAlaaspasna1a 27
Db 167 GTACGAGAGATCTGCTGCC 150

RESULT 26
B1637415/c 567 bp mRNA linear EST 10-SEP-2001
LOCUS SD19430.5prtime SD Drosophila melanogaster schneider L2 cell culture
DEFINITION pot2 Drosophila melanogaster cDNA clone SD19430.5 similar to
CG7289; FBan0007289 GO:[] located on: 2L 22B8-22B8;: 05/19/2001,
mRNA sequence.
ACCESSION B1637415
VERSION B1637415.1 GI:15539625
KEYWORDS EST.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 567)
Harvey, D., Brokstein, P., Hong, L., Evans-Holm, M., Su, C., Tsang, G.,
Lewis, S. and Rubin, G. M.
BDGP/HM1 Drosophila EST project
Unpublished (2001)
Contact: Stapleton, M.

TITLE Lawrence Berkeley National Lab
JOURNAL One Cyclotron Rd, Berkeley, CA 94720, USA
COMMENT Fax: 510 486 6798
Email: http://www.fruitfly.org/EST, east@fruitfly.berkeley.edu
hit genomic AE003584; arm:2L [1824960,2149443]
estimated-cyto:22B8-22D2: 05/19/2001
Plate: SD.194 row: C column: 6
High quality sequence stop: 429.
Location/Qualifiers
1..567
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="SD Drosophila melanogaster schneider L2 cell
culture pot2"
/lab_host="DHS-alpha"
/notes="Vector: pot2; Site_1: EcoRI; Site_2: XhoI; Sized
fractionated cDNAs were directly ligated into pot2.
Plasmid cDNA library."

BASE COUNT 127 a 161 c 141 g 138 t
ORIGIN
Alignment Scores:
Pred. No.: 193
Score: 51.00
Percent Similarity: 53.85%
Best Local Similarity: 42.31%
Query Match: 39.53%
DB: 10

US-09-847-539a-6_COPY_59_86 (1-28) x B1637415 (1-567)
OY 2 AspalaleuGlua1aleuAlaaspGIntThraspalaleuGInserGIua1ala1aVal 21
Db 227 GATCGCGCGAATCATATAGAACACACAGACGAGAGAGAGGAAACAAGCGCG 168
OY 22 VallysAlaaspasna1a 27
Db 167 GTACGAGAGATCTGCTGCC 150

RESULT 27
A1109571/c 568 bp mRNA linear EST 19-APR-2001
LOCUS GH08804.5prtime GH Drosophila melanogaster head pot2 Drosophila
DEFINITION

```

accession melanogaster cDNA clone GH08804 5prime, mRNA sequence.  
 A1109571  
 version A1109571.1 GI:3477895  
 keywords EST.  
 source fruit fly.  
 organism Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 reference 1 (bases 1 to 568)  
 authors Harvey, D., Brokstein, P., Hong, L., Evans-Holm, M., Su, C., Tsang, G.,  
 Lewis, S. and Rubin, G.M.  
 title BDGP/HMT Drosophila EST Project  
 journal Unpublished (2001)  
 comment Contact: Stapleton, M.  
 BDGP  
 Lawrence Berkeley National Lab  
 One Cyclotron Rd, Berkeley, CA 94720, USA  
 Fax: 510 486 6798  
 Email: [http://www.fruitfly.org/EST\\_estefruitfly.berkeley.edu](http://www.fruitfly.org/EST_estefruitfly.berkeley.edu)  
 hit genomic sequence DS06578  
 Plate: 88 row: A column: 4  
 high quality sequence stop: 511.  
 location/Qualifiers  
 1..568  
 /organism="Drosophila melanogaster"  
 /db\_xref="taxon:7227"  
 /clone="GH08804"  
 /clone\_1lb="GH Drosophila melanogaster head port2"  
 /sex="male and female"  
 /dev\_stage="adult"  
 /lab\_host="DH5 - alpha"  
 /note="Organ: head; Vector: pOT2; Site\_1: EcoRI; Site\_2: XhoI; Sized fractionated cDNAs were directly ligated into pOT2. Plasmid cDNA library."  
 BASE COUNT 127 a 161 c 142 g 138 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 193 Length: 568  
 Score: 51.00 Matches: 11  
 Percent Similarity: 53.85% Conservative: 3  
 Best Local Similarity: 42.31% Mismatches: 12  
 Query Match: 39.53% Indels: 0  
 DB: 9 Gaps: 0  
 US-09-847-539a-6\_COPY\_59\_86 (1-28) x A1109571 (1-568)  
 Oy 2 AsphalaleuGluAlaLeuAlaSpGlnThrAspAlaLeuGlnSerGluGluAlaAlaVal 21  
 Db 228 GATCGCGCGAATCATATAGAACACACAGACGACGAGACGAGAACAAACGCGCG 169  
 Oy 22 VallysAlaAspAsnAla 27  
 Db 168 GTACGAGAGATCGTCC 151  
 RESULT 28  
 AA392809 569 bp mRNA linear EST 23-APR-2001  
 LOCUS LD12115.5prime LD Drosophila melanogaster embryo Bluescript  
 DEFINITION Drosophila melanogaster cDNA clone LD12115 5 similar to CG7289;  
 Pspan007289 located on: 2L 22B8-22B8; 04/10/2001, mRNA sequence.  
 AA392809  
 VERSION AA392809.2 GI:13766578  
 KEYWORDS EST.  
 SOURCE fruit fly.  
 ORGANISM Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 reference 1 (bases 1 to 569)  
 authors Harvey, D., Brokstein, P., Hong, L., Evans-Holm, M., Su, C., Tsang, G.,  
 Lewis, S. and Rubin, G.M.

title BDGP/HMT Drosophila EST Project  
 journal Unpublished (2001)  
 comment On Jan 19, 1998 this sequence version replaced gi:2792678.  
 Other\_ESTs: LD12115.3prime  
 Contact: Stapleton, M.  
 BDGP  
 Lawrence Berkeley National Lab  
 One Cyclotron Rd, Berkeley, CA 94720, USA  
 Fax: 510 486 6798  
 Email: [http://www.fruitfly.org/EST\\_estefruitfly.berkeley.edu](http://www.fruitfly.org/EST_estefruitfly.berkeley.edu)  
 hit genomic AE003584; arm:2L [1824960,2149443]  
 estimated-cyto:22B8-22D2; 04/10/2001  
 Plate: LD.121 row: B column: 3  
 high quality sequence stop: 527  
 POLYA-No.  
 location/Qualifiers  
 1..569  
 /organism="Drosophila melanogaster"  
 /db\_xref="BDGP\_EST:BDc1n011306"  
 /db\_xref="taxon:7227"  
 /clone="LD12115"  
 /clone\_1lb="LD Drosophila melanogaster embryo Bluescript"  
 /sex="male and female"  
 /dev\_stage="0 to 24 hours mixed stage embryonic"  
 /lab\_host="SOLR"  
 /note="Organ: embryo; Vector: Bluescript SK; Site\_1: EcoRI  
 ; Site\_2: XhoI; Constructed using Stratagene Zap-cDNA  
 Synthesis kit. Oligo dt-primed and directionally cloned at  
 EcoRI and XhoI in Bluescript SK(+/-)"  
 BASE COUNT 127 a 161 c 143 g 138 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 194 Length: 569  
 Score: 51.00 Matches: 11  
 Percent Similarity: 53.85% Conservative: 3  
 Best Local Similarity: 42.31% Mismatches: 12  
 Query Match: 39.53% Indels: 0  
 DB: 9 Gaps: 0  
 US-09-847-539a-6\_COPY\_59\_86 (1-28) x AA392809 (1-569)  
 Oy 2 AsphalaleuGluAlaLeuAlaSpGlnThrAspAlaLeuGlnSerGluGluAlaAlaVal 21  
 Db 228 GATCGCGCGAATCATATAGAACACACAGACGACGAGACGAGAACAAACGCGCG 169  
 Oy 22 VallysAlaAspAsnAla 27  
 Db 168 GTACGAGAGATCGTCC 151  
 RESULT 29  
 A1388559 581 bp mRNA linear EST 19-APR-2001  
 LOCUS GH19508.5prime GH Drosophila melanogaster head port2 Drosophila  
 DEFINITION melanogaster cDNA clone GH19508 5prime, mRNA sequence.  
 A1388559  
 VERSION A1388559.1 GI:4202570  
 KEYWORDS EST.  
 SOURCE fruit fly.  
 ORGANISM Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 reference 1 (bases 1 to 581)  
 authors Harvey, D., Brokstein, P., Hong, L., Evans-Holm, M., Su, C., Tsang, G.,  
 Lewis, S. and Rubin, G.M.  
 title BDGP/HMT Drosophila EST Project  
 journal Unpublished (2001)  
 comment Contact: Stapleton, M.  
 BDGP  
 Lawrence Berkeley National Lab  
 One Cyclotron Rd, Berkeley, CA 94720, USA  
 Fax: 510 486 6798

Email: [http://www.fruitfly.org/EST\\_est@fruitfly.berkeley.edu](http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu)  
hit genomic sequence AC005749  
plate: 195 row: A column: 8  
High quality sequence stop: 465.  
Location/Qualifiers

## FEATURES

1..581

/organism="Drosophila melanogaster"

/db\_xref="taxon:7227"

/clone\_lib="GH19508"

/clone\_lib="GH Drosophila melanogaster head POT2"

/sex="male and female"

/dev\_stage="adult"

/lab\_host="DHS - alpha"

/note="Organ: head; Vector: POT2; Site\_1: EcoRI; Site\_2: XhoI; Sized fractionated cDNAs were directly ligated into POT2. Plasmid cDNA library."

## BASE COUNT

129 a 164 c 144 g 144 t

## ORIGIN

## Alignment Scores:

Pred. No.: 200 Length: 581  
Score: 51.00 Matches: 11  
Percent Similarity: 53.85 Conservative: 3  
Best Local Similarity: 42.318 Mismatches: 12  
Query Match: 39.538 Indels: 0  
DB: 9 Gaps: 0

US-09-847-539a-6\_copy\_59\_86 (1-28) x A1388559 (1-581)

QY 2 AspAlaLeuCluaLeuAlaAspGlnThrAspAlaLeuCluInserGluCluaAlaAlaVal 21

Db 227 GATCGGGCGGAAATCATATAGAACACACAGACGACAGAGAGAGGAAACAAAGCGCG 168

QY 22 ValIysAlaAspAsnAla 27

Db 167 GTACGAGAGATCGTGC 150

RESULT 30

A1108834/c

LOCUS GH08128.5prfme GH Drosophila melanogaster head POT2 Drosophila

DEFINITION melanogaster cDNA clone GH08128 5prime, mRNA sequence.

ACCESSION A1108834

VERSION A1108834.1 GI:3477369

KEYWORDS EST.

SOURCE fruit fly.

ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 583)

REFERENCE Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,

Lewis,S., and Rubin,G.M.

BDGP/HIMI Drosophila EST Project

Unpublished (2001)

Contact: Stapleton, M.

COMMENT

BDGP Lawrence Berkeley National Lab

One Cyclotron Rd, Berkeley, CA 94720, USA

Fax: 510 486 6798

Email: [http://www.fruitfly.org/EST\\_est@fruitfly.berkeley.edu](http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu)

hit genomic sequence DS06378

plate: 81 row: C column: 4

High quality sequence stop: 492.

Location/Qualifiers

1..583

/organism="Drosophila melanogaster"

/db\_xref="taxon:7227"

/clone\_lib="GH08128"

/clone\_lib="GH Drosophila melanogaster head POT2"

/sex="male and female"

/dev\_stage="adult"

/lab\_host="DHS - alpha"

## BASE COUNT

/note="Organ: head; Vector: POT2; Site\_1: EcoRI; Site\_2: XhoI; Sized fractionated cDNAs were directly ligated into POT2. Plasmid cDNA library."  
129 a 164 c 146 g 144 t

## ORIGIN

1..583

## Alignment Scores:

Pred. No.: 201 Length: 583  
Score: 51.00 Matches: 11  
Percent Similarity: 53.85 Conservative: 3  
Best Local Similarity: 42.318 Mismatches: 12  
Query Match: 39.538 Indels: 0  
DB: 9 Gaps: 0

US-09-847-539a-6\_copy\_59\_86 (1-28) x A1108834 (1-583)

QY 2 AspAlaLeuCluaLeuAlaAspGlnThrAspAlaLeuCluInserGluCluaAlaAlaVal 21

Db 228 GATCGGGCGGAAATCATATAGAACACACAGACGACAGAGAGAGGAAACAAAGCGCG 169

QY 22 ValIysAlaAspAsnAla 27

Db 168 GTACGAGAGATCGTGC 151

RESULT 31

A2698116/c

LOCUS RPCI-23-224F10.TV RPCI-23 Mus musculus genomic clone RPCI-23-224F10

DEFINITION ' DNA sequence.

A2698116

VERSION A2698116.1 GI:12414882

KEYWORDS GSS.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 615)

REFERENCE Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Aklnet

,B., Levins,M., Moggan,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.

and Fraser,C.M.

Mouse BAC End Sequences from Library RPCI-23

Unpublished (1999)

Contact: Shaying Zhao

Other-GSS: RPCI-23-224F10.TJ

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: [szhao@tigr.org](mailto:szhao@tigr.org)

Clones are derived from the mouse BAC library RPCI-23. For BAC

library availability, please contact Pieter de Jong

(pdejong@mail.cho.org). Clones may be purchased from BACPAC

Resources (<http://www.choi.org/bacpac/orderingframe.htm>). BAC end

plate: 224 row: F column: 10

Seq primer: T7

Class: BAC ends.

Location/Qualifiers

1..615

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone\_lib="RPCI-23-224F10"

/clone\_lib="RPCI-23"

/sex="female"

/lab\_host="DH10B"

/note="Organ: kidney/Brain; Vector: pBAC3.6; Site\_1: EcoRI; Site\_2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBAC3.6 vector at the EcoRI sites. The ligation products were transformed into

DH10B electrocompetent cells (BRL Life Technologies).

BASE COUNT	1
ORIGIN	

DH10B electrocompetent cells (BRL Life Technologies)

Pred. No.:	216	Length:	615
Score:	51.00	Matches:	11

Percent Similarity:	68.18%	Conservative:	4
Best Local Similarity:	50.00%	Mismatches:	7
Conservative:	50.00%	Mismatches:	7

```

Query match: 39.336 Inserts: 0
DB: 12 Gaps: 0

```

US-09-847-539A-6\_COPY\_59\_86 (1-28) x AZ698116 (1-615)

QY 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSer

Db 522 TCAGATCCTCTGGAGAGACTTAAGAATCAGACAGATTCCCTAAAAGC

0y	21	val	22

Db 462 GTAGTG 457

RESULT 32  
BB201301/c

LOCUS	622 bp	mRNA	11
DEFINITION	BB201301	RIKEN full-length enriched, 0 day ne	

musculus	CDNA	clone A430036B19	3', mRNA sequence
BB201301			
ACCESSION			
BB201301			

VERSION BB201301.2 GI:162/1557  
KEYWORDS EST.  
SOURCE

SOURCE	ORGANISM	Substrate	Character	Condition	Notes
house mouse.	Mus musculus				

REFERENCE  
1. (basos 1 to 523)  
Mammalia; Eutheria; Rodentia; Sciurognathi; N  
Eukaryota; Metazoa; Chordata; Verte  
Eukaryota; Metazoa; Chordata; Verte

**REFERENCE**  
1. (cases 1 to 622)

**AUTHORS**  
Arakawa, T., Carninci, P., Fukuda, S., Furuno, M.  
Hiramoto, K., Hori, F., Ishii, Y., Ito, M., Kawano,  
K., Kohno, T., Kubota, Y., Kurita, A., Matsuda, N.,  
Miyazaki, T., Nakamura, T., Nakayama, M., Ohtsuka,

M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, Y., Okada, T., Saito, R., Sakai, C., Sakai, M., Tanaka, H., Takahashi, T., Terakura, Y., Ueda, N., Yamashita, T., Yoshida, T., Yokoyama, C., Yokoyama, H., Yokoyama, T., Yokoyama, Y., Yokoyama, Z., Yokoyama, A., Yokoyama, B., Yokoyama, C., Yokoyama, D., Yokoyama, E., Yokoyama, F., Yokoyama, G., Yokoyama, H., Yokoyama, I., Yokoyama, J., Yokoyama, K., Yokoyama, L., Yokoyama, M., Yokoyama, N., Yokoyama, O., Yokoyama, P., Yokoyama, Q., Yokoyama, R., Yokoyama, S., Yokoyama, T., Yokoyama, U., Yokoyama, V., Yokoyama, W., Yokoyama, X., Yokoyama, Y., Yokoyama, Z.

D., Shibata, K., Shinagawa, A., Shiraki, T., So-  
Tadami, M., Tadawa, A., Takahashi, F., Takeda, Y.

TITLE  
Muramatsu, M. and Hayashizaki, Y.  
RIKEN Mouse ESTs (Arakawa, T., et al. 2001)

JOURNAL Unpublished (2001)  
COMMENT On Jun 30, 2000 this sequence version replaced

Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group

Sciences Center(GSC), Yokohama Institute  
The Institute of Physical and Chemical Research

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kan  
Tel: 81-45-503-9222

Fax: 81-45-503-9216  
Email: genome-res@sc.riken.go.jp,  
and h45@sc.riken.go.jp

URL:

'M., Kounio, H., Okazaki, I., Muramatsu, M. and  
Normalization and subtraction of cap-trapped  
prepare full-length cDNA libraries for rapid

Genome Res. • 10 (10), 1617-1630 (2000)

Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K.  
S. Kawai, T. Okazaki, Y. Muramatsu, M. Inoue,  
Hagi, N., Fujimake, S., Inoue, N., Ogawa, I.,  
Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K.

RIKEN Integrated sequence analysis (RISA) system  
Hayashizaki, Y.

sequencing pipeline with 384 multicapillary s  
10 (11), 1757-1771 (2000)

Konno, H., Fukunishi, Y., Shibata, K., Itoh, M.,  
, Y. and Hayashizaki, Y.

Computer-based methods for the mouse full-length  
encyclopedia: real-time sequence clustering

nonredundant cDNA library: Genome Res. 11  
Kondo, S., Shinagawa, A., Saito, T., Kiyosawa, H.

, K., Fukuda, S., Hara, A., Itoh, M., Kawai, J.,

TITLE Lewis, S. and Rubin, G.M.  
JOURNAL BDGP/HIMI Drosophila EST Project  
COMMENT Unpublished (2001)  
Contact: Stapleton, M.  
BDGP

Lawrence Berkeley National Lab  
One Cyclotron Rd, Berkeley, CA 94720, USA  
Fax: 510 486 6798  
Email: [http://www.fruitfly.org/EST\\_est@fruitfly.berkeley.edu](http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu)  
hit genomic sequence DS06378  
Plate: 21 row: G column: 1  
High quality sequence stop: 567.  
Location/Qualifiers

## FEATURES

source

1. 633

/organism="Drosophila melanogaster"

/db\_xref="taxon:7227"

/clone="GH02173"

/clone\_lib="GH Drosophila melanogaster head POT2"

/sex="male and female"

/dev\_stage="adult"

/lab\_host="DH5 - alpha"

/note="Organ: head; Vector: POT2; Site: 1; EcoRI: Site 2; XhoI: sized fractionated cDNAs were directly ligated into POT2 plasmid cDNA library."

BASE COUNT 140 a 182 c 153 g 158 t

ORIGIN

## Alignment Scores:

Pred. No.: 225

Length: 633

Score: 51.00

Matches: 11

Percent Similarity: 53.85%

Conservative: 3

Best Local Similarity: 42.31%

Mismatches: 12

Query Match: 39.53%

Indels: 0

Gaps: 0

US-09-847-539a-6\_COPY\_59\_86 (1-28) x AI062828 (1-633)

QY 2 ASPALALEUGLUALALEUALASPGLNTHASPALALEUGINSERGIUGLUALALEVAL 21

DB 227 GATCGGGCGAATCATATAGAACACACAGACGACGAGGAGGAACAAGCGCGC 168

QY 22 VALLYSALASPASNAIA 27

DB 167 GTACGAGAGATCGTGCC 150

RESULT 34

AA391497/c

LOCUS 643 bp mRNA linear EST 19-APR-2001

DEFINITION LD10172.5prime LD Drosophila melanogaster embryo Bluescript

ACCESSION AA391497

VERSION AA391497.1 GI:2044472

KEYWORDS EST.

SOURCE fruit fly.

ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 643)

REFERENCE Harvey, D., Brokslein, P., Hong, L., Evans-Holm, M., Su, C., Tsang, G.,

LEWIS, S. and RUBIN, G.M.

AUTHORS BDGP/HIMI Drosophila EST Project

JOURNAL Unpublished (2001)

COMMENT Contact: Stapleton, M.  
BDGP

Lawrence Berkeley National Lab

One Cyclotron Rd, Berkeley, CA 94720, USA

Fax: 510 486 6798

Email: [http://www.fruitfly.org/EST\\_est@fruitfly.berkeley.edu](http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu)

Plate: 101 row: F column: 12

High quality sequence stop: 463.

Location/Qualifiers

1. 643

source

/organism="Drosophila melanogaster"

/db\_xref="BDGP\_EST:BDCLN009443"

/clone="LD10172"

/clone\_lib="LD Drosophila melanogaster embryo Bluescript"

/sex="male and female"

/dev\_stage="0 to 24 hours mixed stage embryonic"

/lab\_host="SOLR"

/note="Organ: embryo; Vector: Bluescript SK; Site: 1; EcoRI: Site 2; XhoI: constructed using Stratagene ZAP-cDNA synthesis kit. Oligo dt-primed and directionally cloned at EcoRI and XhoI in Bluescript SK(+/-)"

BASE COUNT 140 a 185 c 154 g 164 t

ORIGIN

## Alignment Scores:

Pred. No.: 230

Length: 643

Score: 51.00

Matches: 11

Percent Similarity: 53.85%

Best Local Similarity: 42.31%

Query Match: 39.53%

Mismatches: 12

Indels: 0

Gaps: 0

US-09-847-539a-6\_COPY\_59\_86 (1-28) x AA391497 (1-643)

QY 2 ASPALALEUGLUALALEUALASPGLNTHASPALALEUGINSERGIUGLUALALEVAL 21

DB 213 GATCGGGCGAATCATATAGAACACACAGACGACGAGGAGGAACAAGCGCGC 154

QY 22 VALLYSALASPASNAIA 27

DB 153 GTACGAGAGATCGTGCC 136

RESULT 35

BE337641

LOCUS 650 bp mRNA linear EST 14-JUL-2000

DEFINITION 894047D11.Y1 C. reinhardtii CC-1690, normalized, lambda zap II

ACCESSION BE337641

VERSION BE337641.1 GI:9210726

KEYWORDS EST.

SOURCE Chlamydomonas reinhardtii.

ORGANISM Chlamydomonas reinhardtii

Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;

Chlamydomonadaceae; Chlamydomonas.

1 (bases 1 to 650)

REFERENCE Grossman, A., Davies, J., Federspiel, N., Harris, E., Lefebvre, P.,

McDermott, V.P., Sillow, C., Stern, D. and Surzycki, R.,

Analyses of the Chlamydomonas reinhardtii Genome: A Model,

Unicellular System for Analyzing Gene Function and Regulation in

Vascular Plants: project phase 2

JOURNAL Unpublished (2000)

COMMENT Contact: Elizabeth H. Harris

DCMB Box 9100

Duke University

Durham, NC 27708-1000, USA

Tel: 919 613 8164

Fax: 919 613 8177

Email: [chlamy@duke.edu](mailto:chlamy@duke.edu).

Location/Qualifiers

1. 650

/organism="Chlamydomonas reinhardtii"

/strain="CC-1690 wild type mt+ 21gr"

/clone\_lib="C. reinhardtii CC-1690, normalized, lambda zap II"

/note="Vector: pBluescript II SK-; Site: 1; EcoRI: Site 2; XhoI: This library, constructed by John Davies and Jeffrey

McDermott, combines cDNAs from CC-1690 cells grown to

mid-log phase in TAP (acetate-containing) medium in the

light, TAP medium in the dark, HS (minimal) medium in

ambient levels of CO2 and HS medium bubbled with 5% CO2.

PolyA mRNA was purified from each sample, pooled and cDNA

synthesized. The cDNA was directionally cloned into lambda ZAP II (Stratagene) in the EcoRI (5') and XhoI (3') sites. plasmidscript II SK- plasmids were excised from the lambda ZAP clones by superinfection with ExAssist (Stratagene) phage. The library was normalized using method 4 described in Bonaldo et al (1996) Genome Research 6: 791-806."

BASE COUNT 153 a 161 c 206 g 130 t

ORIGIN

Alignment Scores:

Pred. No.:	233	Length:	650
Score:	51.00	Matches:	11
Percent Similarity:	62.96%	Conservative:	6
Best Local Similarity:	40.74%	Mismatches:	10
Query Match:	39.53%	Indels:	0
DB:	10	Gaps:	0

US-09-847-539a-6\_COPY\_59\_86 (1-28) x BE337641 (1-650)

OY 2 AspalaleuGluaAlaAspGlnThrAspalaleuGlnSerGluAlaAlaVal 21

Db 507 GAGAGATCGAGCTGCTGTCGACACGACGCTGCGCACACGCGCGGACGCTT 566

OY 22 ValysAlaAspaAlaAla 28

Db 567 CCAAAAGAGGAGCGGACGCT 587

RESULT 36

BI633216/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BDGP

FEATURES

source

BASE COUNT

ORIGIN

Alignment Scores:

Pred. No.:

238

Length:

660

Score:	51.00	Matches:	11
Percent Similarity:	53.85%	Conservative:	3
Best Local Similarity:	42.31%	Mismatches:	12
Query Match:	39.53%	Indels:	0
DB:	10	Gaps:	0

US-09-847-539a-6\_COPY\_59\_86 (1-28) x BI633216 (1-660)

OY 2 AspalaleuGluaAlaAspGlnThrAspalaleuGlnSerGluAlaAlaVal 21

Db 226 GATCGGCGGAATCATATAGAGACACACAGCAGAGGAGGAGGAAAGCGCG 167

OY 22 ValysAlaAspaAla 27

Db 166 GTACGAGGAGATCTGCC 149

RESULT 37

AI062753/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BDGP

FEATURES

source

BASE COUNT

ORIGIN

Alignment Scores:

Pred. No.:

Score:

Percent Similarity:

Best Local Similarity:

Query Match:

DB:

US-09-847-539a-6\_COPY\_59\_86 (1-28) x AI062753 (1-666)

OY 2 AspalaleuGluaAlaAspGlnThrAspalaleuGlnSerGluAlaAlaVal 21

Db 227 GATCGGCGGAATCATATAGAGACACACAGCAGAGGAGGAGGAAAGCGCG 168

OY 22 ValysAlaAspaAla 27

||||: ||| |||  
Db 167 GTACGAGAGATCGTCC 150

RESULT 38  
LOCUS A2699109/c 666 bp DNA linear GSS 24-JAN-2001  
DEFINITION RPCI-23-224J19.TV RPCI-23 Mus musculus genomic clone RPCI-23-224J19  
ACCESSION A2699109  
VERSION A2699109  
KEYWORDS GSS  
SOURCE house mouse.  
ORGANISM Mus musculus  
REFERENCE Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Eukaryota; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus;  
1 (bases 1 to 666)  
AUTHORS Zhao, S., Nierman, W., Feldblyum, T., Malek, J., Shatsman, S., Alinret,  
B., Levins, M., McGinn, S., Tsengaye, G., Geer, K., Krol, M., de Jong, P.,  
and Fraser, C. M.  
TITLE Mouse BAC End Sequences from Library RPCI-23  
JOURNAL Unpublished (1999)  
COMMENT Other GSS: RPCI-23-224J19.TJ  
Contact: Shaying Zhao  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: szhao@tigr.org  
Clones are derived from the mouse BAC library RPCI-23. For BAC  
library availability, please contact Pieter de Jong  
(pdejong@mail.cho.org). Clones may be purchased from BACPAC  
Resources (<http://www.choi.org/bacpac/orderingframe.html>). BAC end  
Page: [http://www.tigr.org/cdb/bac\\_ends/mouse/bac\\_end\\_intro.html](http://www.tigr.org/cdb/bac_ends/mouse/bac_end_intro.html)  
Plate: 224 row: J column: 19  
Seq primer: T7  
Classes: BAC ends.  
Location/Qualifiers  
1..666  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="RPCI-23-224J19"  
/clone\_lib="RPCI-23"  
/sex="Female"  
/lab\_host="DH10B"  
/note="Organ: Kidney/Brain; Vector: pBAC3.6; Site\_1:  
EcoRI; Site\_2: EcoRI; Female C57BL/6J mouse kidney and/or  
brain genomic DNA was isolated and partially digested  
with a combination of EcoRI and EcoRI Methylase. Size  
selected DNA was cloned into the pBAC3.6 vector at the  
EcoRI sites. The ligation products were transformed into  
DH10B electrocompetent cells (BRL Life Technologies)."  
BASE COUNT 203 a 163 c 150 g 150 t  
ORIGIN

Alignment Scores:  
Pred. No.: 241 Length: 666  
Score: 51.00 Matches: 11  
Percent Similarity: 68.18% Conservative: 4  
Best Local Similarity: 50.00% Mismatches: 7  
Query Match: 39.53% Indels: 0  
DB: 12 Gaps: 0

US-09-847-539a-6\_copy\_59\_86 (1-28) x A2699109 (1-666)  
Qy 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAla 20  
Db 522 TCAGATCCCTCGGAGAGACTTAAGATACACAGACATTCCCTAAAGGACAGACAGAGAG 463  
Qy 21 ValVal 22  
Db 462 GTAGTG 457

RESULT 39  
LOCUS B1637204/c 710 bp mRNA linear EST 10-SEP-2001  
DEFINITION SD19151.5prime SD Drosophila melanogaster Schneider L2 cell culture  
PORT2 Drosophila melanogaster cDNA clone SD19151 5 similar to  
CG7289: FBan0007289 GO:[] located on: 2L 22B8-22B8:: 05/19/2001,  
mRNA sequence.  
ACCESSION B1637204  
VERSION B1637204  
KEYWORDS EST  
SOURCE fruit fly.  
ORGANISM Drosophila melanogaster  
REFERENCE Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 710)  
AUTHORS Harvey, D., Brokstein, P., Hong, L., Evans-Holm, M., Su, C., Tsang, G.,  
Lewis, S., and Rubin, G. M.  
TITLE BDCP/HMI Drosophila EST Project  
JOURNAL Unpublished (2001)  
COMMENT Contact: Stapleton, M.  
BDCP  
Lawrence Berkeley National Lab  
One Cyclotron Rd, Berkeley, CA 94720, USA  
Fax: 510 486 6798  
Email: <http://www.fruitfly.org/EST/estfruitfly.berkeley.edu>  
hit genomic AE003584: arm:2L [1824960,2149443]  
estimated-cyto:22B4-22D2: 05/19/2001  
Plate: SD.191 row: E column: 3  
High quality sequence sloop: 578.  
Location/Qualifiers  
1..710  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/clone="SD19151"  
/clone\_lib="SD Drosophila melanogaster Schneider L2 cell  
culture PORT2"  
/lab\_host="DH5-alpha"  
/note="Vector: PORT2; Site\_1: EcoRI; Site\_2: XhoI; Sized  
fractionated cDNAs were directly ligated into PORT2.  
Plasmid cDNA library."  
BASE COUNT 165 a 203 c 170 g 171 t 1 others  
ORIGIN

Alignment Scores:  
Pred. No.: 264 Length: 710  
Score: 51.00 Matches: 11  
Percent Similarity: 53.85% Conservative: 3  
Best Local Similarity: 42.31% Mismatches: 12  
Query Match: 39.53% Indels: 0  
DB: 10 Gaps: 0

US-09-847-539a-6\_copy\_59\_86 (1-28) x B1637204 (1-710)  
Qy 2 AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaVal 21  
Db 228 GATCGCGCGAATCATATTAAGACACACACAGACGACGAGACGAGACGAGAACAGCGCG 169  
Qy 22 ValLyAspAlaAspAla 27  
Db 168 GTACGAGAGATCGTCC 151

RESULT 40  
LOCUS BF494445/c 797 bp mRNA linear EST 19-APR-2001  
DEFINITION AT03046.5prime AT Drosophila melanogaster adult testes PORT7  
Drosophila melanogaster cDNA clone AT03046 5 similar to CG7289:  
FBan0007289 located on: 2L 22B8-22B8:: 04/07/2001, mRNA sequence.  
ACCESSION BF494445  
VERSION BF494445  
KEYWORDS EST.  
SOURCE fruit fly.



ORGANISM *Drosophila melanogaster*

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

## REFERENCE

1 (bases 1 to 797)

Stapleton,M., Brokstein,P., Hong,L., Agbayani,A., Baxter,E., Berman  
 ,B., Carlson,J., Champe,M., Chavez,C., Chew,M., Dorsett,V., Farfan  
 ,D., Frise,E., George,R., Gonzalez,M., Guadin,H., Harris,N., Li,P.,  
 Liao,G., Miranda,A., Mista,S., Mungall,C.J., Nunoo,J., Pacleb,J.,  
 Paragas,V., Park,S., Phouanavong,S., Wan,K., Yu,C., Lewis,S.E.,  
 Celisner,S. and Rubin,G.M.

## TITLE

## JOURNAL

## COMMENT

BDGP/HMI AT *Drosophila* EST Project  
 Unpublished (2000)  
 On Dec 6, 2000 this sequence version replaced gi:11577746.  
 Contact: Stapleton, M.

BDGP  
 Lawrence Berkeley National Lab  
 One Cyclotron Rd, Berkeley, CA 94720, USA  
 Fax: 510 486 6798  
 Email: [http://www.fruitfly.org/EST\\_estefruitfly.berkeley.edu](http://www.fruitfly.org/EST_estefruitfly.berkeley.edu)  
 hit genomic AE003584: arm:2L [1824960,2149443]  
 estimated-cyto:22B4-22D2: 04/07/2001  
 Plate: AT.30 row: D column: 10  
 High quality sequence stop: 699.  
 Location/Qualifiers

## FEATURES

## source

1..797  
 /organism="Drosophila melanogaster"  
 /db\_xref="taxon:7227"  
 /clone\_lib="AT Drosophila melanogaster adult testes potb7"  
 /sex="male"  
 /dev\_stage="0-3 day old Ore-R males" )  
 /lab\_host="Plates AT.10-AT.120: DHS-alpha. Plates  
 AT.121-AT.319: DHS-alpha Tona"  
 /note="Organ: ADULT testes; Vector: potb7; Site:1: EcoRI;  
 Site:2: XhoI; The mRNA for the testis library was made  
 from testes and seminal vesicles hand dissected from 0-3  
 day old Ore-R males. RNA kindly provided by the lab of  
 Margaret Fuller. Sized fractionated cDNAs were directly  
 ligated into potb7. Plasmid cDNA library."

BASE COUNT 178 a 234 c 195 g 190 t  
 ORIGIN

## Alignment Scores:

Pred. No.:	310	Length:	797
Score:	51.00	Matches:	11
Percent Similarity:	53.85%	Conservative:	3
Best Local Similarity:	42.31%	Mismatches:	12
Query Match:	39.53%	Indels:	0
DB:	10	Gaps:	0

US-09-847-539a-6\_COPY\_59\_86 (1-28) x BP494445 (1-797)

QY 2 AsPALaLeuglAaLeuAlaAspGlnThrAsPALaLeuglInserGlnGluAlaAlaVal 21  
 ||| ::||| ||||| |||  
 Db 203 GATGGCGGAATCATATAGAACACACAGACGACAGAGAGAGAGAAACAGCGCG 144  
 |||::||| |||  
 QY 22 VallysaLaaspaNaLa 27  
 |||::||| |||  
 Db 143 GTACGAGAGATCGTCC 126

Search completed: October 13, 2002, 06:10:16  
 Job time : 333.61 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2002, 03:46:02 : Search time 12.1283 Seconds  
(without alignments)  
399.383 Million cell updates/sec

Title: US-09-847-539A-6\_COPY\_59\_86  
Perfect score: 129  
Sequence: 1 SDALFALADQTDALQSEEAAYVKNADNA 28

Scoring table: BLOSUM62  
Gapop 10.0 , Gapept 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

SPTREMBL\_19:\*  
1: sp.archaea:\*  
2: sp.bacteria:\*  
3: sp.fungi:\*  
4: sp.human:\*  
5: sp.invertebrate:\*  
6: sp.mammal:\*  
7: sp.mhc:\*  
8: sp.organelle:\*  
9: sp.phage:\*  
10: sp.plant:\*  
11: sp.podent:\*  
12: sp.virus:\*  
13: sp.vertibrate:\*  
14: sp.unclassified:\*  
15: sp.virus:\*  
16: sp.bacteriap:\*  
17: sp.archeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	129	100.0	156	2	Q9S6G3 streptococc
2	129	100.0	156	2	Q9S6G4 streptococc
3	129	100.0	217	16	Q9X5C4 streptococc
4	129	100.0	239	2	Q9X5C5 streptococc
5	129	100.0	268	2	Q9X5C6 streptococc
6	57	44.2	377	5	076220 tetrahymena
7	57	44.2	412	16	025885 helicobacte
8	57	44.2	412	16	092507 helicobacte
9	57	44.2	850	2	Q935B1 salmoneilla
10	51.5	39.5	999	2	Q931X0 streptomyc
11	51	39.5	203	16	Q92PR3 listeria in
12	51	39.5	276	2	086553 streptomyc
13	51	39.5	486	17	Q9HNV3 halobacteri
14	51	39.5	4589	5	076506 tetrahymena
15	50	38.8	1968	5	020439 caenorhabdi
16	49.5	38.4	396	2	087677 amycolatops

17	49	38.0	207	2	Q9JP85 streptomyc
18	49	38.0	251	16	Q92P10 rhizobium m
19	49	38.0	445	10	Q9SVK1 arabidopsis
20	49	38.0	684	17	Q9HPR2 halobacteri
21	48	37.2	207	9	Q94MN6 bacterioph
22	48	37.2	1160	10	Q9FHF8 arabidopsis
23	47	36.4	254	16	084403 chlamydia t
24	47	36.4	254	16	Q9PJ29 chlamydia m
25	47	36.4	387	2	Q9KXW4 streptomyc
26	47	36.4	1617	5	Q931X4 leishmania
27	46.5	36.0	536	17	Q9HPA2 halobacteri
28	46.5	36.0	541	17	Q979W1 thermoplasm
29	46.5	36.0	749	2	Q9KZ05 streptomyc
30	46.5	36.0	784	2	086858 streptomyc
31	46	35.7	106	16	Q9K0Z7 neisseria m
32	46	35.7	208	2	082964 burkholderi
33	46	35.7	220	2	Q9K4B4 streptomyc
34	46	35.7	237	2	Q9AD77 streptomyc
35	46	35.7	250	2	Q9ZAK1 streptomyc
36	46	35.7	299	10	Q9FCP2 arabidopsis
37	46	35.7	303	10	Q9LUF2 arabidopsis
38	46	35.7	341	12	Q91GV4 sagliyama v1
39	46	35.7	345	16	Q9RQV2 caulobacter
40	46	35.7	351	12	Q9E338 alphavirus
41	46	35.7	353	16	Q9WY02 thermotoga
42	46	35.7	367	2	Q91034 streptomyc
43	46	35.7	526	16	Q91491 pseudomonas
44	46	35.7	539	16	Q9KTB8 bacillus ha
45	46	35.7	732	17	Q9HMP9 halobacteri

#### ALIGNMENTS

RESULT 1  
ID Q9S6G3 PRELIMINARY; PRT; 156 AA.

AC Q9S6G3: 09S6G3: 1  
DT 01-MAY-2000 (TREMUREL. 13, Created)  
DT 01-MAY-2000 (TREMUREL. 13, Last sequence update)  
DT 01-DEC-2001 (TREMUREL. 19, Last annotation update)  
DE GRAB (FRAGMENT).  
OS Streptococcus pyogenes.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID-1314;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-AP1;  
RX MEDLINE-99269061; PubMed-10336419;  
RA Rasmussen M., Muller H.P., Bjorck L.;  
RT bacterial surface by binding alpha2-macroglobulin.";  
RL J. Biol. Chem. 274:15336-15344(1999).  
DR EMBL: AF124400; AD26339.1; -;  
DR InterPro: IPR001899; Gram\_pos\_anchor.  
DR PROSITE: PS00343; GRAM\_POS\_ANCHORING; UNKNOWN\_1.  
FT NON\_TER  
FT NON\_TER  
SQ SEQUENCE 156 AA; 16587 MW; 19FAA98D0599D866 CRC64;

Query Match 100.0%; Score 129; DB 2; Length 156;  
Best Local Similarity 100.0%; Pred. No. 6.5e-10;  
Matches 28; Conservative 0; Mismatches 0; Indels 0;

QY 1 SDALFALADQTDALQSEEAAYVKNADNA 28  
|||||  
Db 59 SDALFALADQTDALQSEEAAYVKNADNA 86

RESULT 2  
Q9S6G4 PRELIMINARY; PRT; 156 AA.

```

AC 09S6G4;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE GRAB (FRAGMENT).
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-KTL3;
RX MEDLINE=9269061; PubMed=10336419;
RA Rasmussen M., Muller H.P., Bjorck L.;
RT *Protein GRAB of streptococcus pyogenes regulates proteolysis at the
RL bacterial surface by binding alpha2-macroglobulin.";
DR EMBL; AF124401; AAD26340.1; -.
DR InterPro: IPR001899; Gram_pos_anchor.
DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
FT NON_TER 1 156
FT NON_TER 156 156
SQ SEQUENCE 156 AA; 16575 MW; 0C20967E2FB0D866 CRC64;

Query Match . 100.0%; Score 129; DB 2; Length 156;
Best Local Similarity 100.0%; Pred. No. 6.5e-10;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SDALEALADQTDALQSEEAAYVKADNAA 28
Db 59 SDALEALADQTDALQSEEAAYVKADNAA 86

RESULT 3
OYX5C4 PRELIMINARY; PRT; 217 AA.
ID 09X5C4;
AC 09X5C4;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE GRAB PRECURSOR (PROTEIN GRAB) (PROTEIN G-RELATED ALPHA 2M-BINDING
DE PROTEIN).
GN GRAB OR SPY1357.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC700294;
RA Rasmussen M., Muller H.P., Bjorck L.;
RT *Protein GRAB of streptococcus pyogenes regulates proteolysis at the
RL bacterial surface by binding alpha2-macroglobulin.";
RT Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-SF370 / ATCC 700294 / SEROTYPE M1;
RX MEDLINE=21192684; PubMed=11296296;
RA Ferretti J.J., Mcshan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
RA Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,
RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,
RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
RT *Complete genome sequence of an M1 strain of Streptococcus pyogenes.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
DR EMBL; AF124399; AAD26338.1; -.
DR EMBL; AE006573; AKK34185.1; -.
DR InterPro: IPR001899; Gram_pos_anchor.
DR Pfam: PF00746; Gram_pos_anchor; 1.
DR PRINTS: PR00015; GP0SANCHOR.
DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
KW Complete proteome; signal; Transmembrane.
FT SIGNAL 1 33
FT SIGNAL 33 33
FT CHAIN 34 217 GRAB.
SQ

```

```

SQ SEQUENCE 217 AA; 22836 MW; 79AA8C4FF5F3FA06 CRC64;

Query Match . 100.0%; Score 129; DB 16; Length 217;
Best Local Similarity 100.0%; Pred. No. 9.4e-10;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SDALEALADQTDALQSEEAAYVKADNAA 28
Db 92 SDALEALADQTDALQSEEAAYVKADNAA 119

RESULT 4
OYX5C5 PRELIMINARY; PRT; 239 AA.
ID 09X5C5;
AC 09X5C5;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE GRAB (FRAGMENT).
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-KTL9;
RX MEDLINE=9269061; PubMed=10336419;
RA Rasmussen M., Muller H.P., Bjorck L.;
RT *Protein GRAB of streptococcus pyogenes regulates proteolysis at the
RL bacterial surface by binding alpha2-macroglobulin.";
RL J. Biol. Chem. 274:15336-15344(1999).
DR EMBL; AF124402; AAD26341.1; -.
DR InterPro: IPR001899; Gram_pos_anchor.
DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
FT NON_TER 1 239
FT NON_TER 239 239
SQ SEQUENCE 239 AA; 24846 MW; 61AC4F6F863AF0F5 CRC64;

Query Match . 100.0%; Score 129; DB 2; Length 239;
Best Local Similarity 100.0%; Pred. No. 1e-09;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SDALEALADQTDALQSEEAAYVKADNAA 28
Db 87 SDALEALADQTDALQSEEAAYVKADNAA 114

RESULT 5
OYX5C6 PRELIMINARY; PRT; 268 AA.
ID 09X5C6;
AC 09X5C6;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE GRAB (FRAGMENT).
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-AP49;
RX MEDLINE=9269061; PubMed=10336419;
RA Rasmussen M., Muller H.P., Bjorck L.;
RT *Protein GRAB of streptococcus pyogenes regulates proteolysis at the
RL bacterial surface by binding alpha2-macroglobulin.";
RL J. Biol. Chem. 274:15336-15344(1999).
DR EMBL; AF124403; AAD26342.1; -.
DR InterPro: IPR001899; Gram_pos_anchor.
DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
FT NON_TER 1 268
FT NON_TER 268 268
SQ SEQUENCE 268 AA; 27744 MW; 1C59239260CDC7E7 CRC64;

```

Query Match 100.0%; Score 129; DB 2; Length 268;  
Best Local Similarity 100.0%; Pred. No. 1.2e-09;  
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SDALADQTDALQSEEAAYVKADNNA 28  
|||||  
DB 171 SDALADQTDALQSEEAAYVKADNNA 198

## RESULT 6

ID 076220 PRELIMINARY; PRT; 377 AA.  
AC 076220;  
DT 01-NOV-1998 (TREMBlrel. 08, Created)  
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE GRANULE LATTICE PROTEIN 5.  
GN GRL5.  
OS Tetrathymena thermophila.  
OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;  
OC Tetrathymenina; Tetrathymena.  
OX NCBI\_TaxID=5911;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-B1686;  
RA Verbsky J.W., Turkewitz A.P.;  
RT "Proteolytic processing and Ca2+ binding activity of dense-core  
RT vesicle polypeptides in Tetrathymena.";  
RL Mol. Biol. Cell 0:0-0(1997).  
DR EMBL: AF031321; AAC27989.1; -  
FT CHAIN 58 377  
SQ SEQUENCE 377 AA; 41509 MW; 67FE32D8035D15C CRC64;

Query Match 44.2%; Score 57; DB 5; Length 377;  
Best Local Similarity 50.0%; Pred. No. 8.8;  
Matches 12; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

OY 2 DALEADQTDALQSEEAAYVKAD 25  
|||  
DB 133 DAQALADNTDALNAEEALANAE 156

## RESULT 7

ID 025885 PRELIMINARY; PRT; 412 AA.  
AC 025885;  
DT 01-JAN-1998 (TREMBlrel. 05, Created)  
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE HYPOTHETICAL 47.1 KDA PROTEIN.  
GN HPI327.  
OS Helicobacter pylori (Campylobacter pylori).  
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;  
OC Helicobacter.  
OX NCBI\_TaxID=210;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-26695 / ATCC 700392;  
RX MEDLINE=97394467; PubMed=9252185;  
RA Tomb J.-F., White O., Kerslavage A.R., Clayton R.A., Sutton G.G.,  
RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,  
RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,  
RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,  
RA McGenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,  
RA Berg D.E., Gocayne J.D., Ullrich T.R., Peterson J.D., Kelley J.M.,  
RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Wathey L., Wallin E.,  
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,  
RA Venter J.C.;  
RT "The complete genome sequence of the gastric pathogen Helicobacter  
RT pylori.";  
RL Nature 388:539-547(1997).  
DR EMBL: AE000634; AAD08377.1; -

DR TIGR: HPI327; -  
DR InterPro: IPR003423; OEP.  
DR Pfam: PF02321; OEP; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 412 AA; 47115 MW; 1724D84A9BPA15DD CRC64;

Query Match 44.2%; Score 57; DB 16; Length 412;  
Best Local Similarity 52.2%; Pred. No. 9.7;  
Matches 12; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

OY 4 LEALADQTDALQSEEAAYVKADN 26  
::|||  
DB 49 IQALQEQIDALSSQEKVSKMDN 71

## RESULT 8

ID 092J07 PRELIMINARY; PRT; 412 AA.  
AC 092J07;  
DT 01-MAY-1999 (TREMBlrel. 10, Created)  
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE PUTATIVE.  
GN JHP1247.  
OS Helicobacter pylori j99 (Campylobacter pylori j99).  
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;  
OC Helicobacter.  
OX NCBI\_TaxID=85963;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99120557; PubMed=9923682;  
RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,  
RA Smith D.R., Noonan B., Guild B.C., deJonge B.V., Carmel G.,  
RA Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,  
RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,  
RA Trust T.J.;  
RT "Genomic sequence comparison of two unrelated isolates of the human  
RT gastric pathogen Helicobacter pylori.";  
RL Nature 397:176-180(1999).  
DR EMBL: AE001548; AAD06820.1; -  
DR InterPro: IPR003423; OEP.  
DR Pfam: PF02321; OEP; 1.  
KW Complete proteome.  
SQ SEQUENCE 412 AA; 47054 MW; 20E1E2904869378F CRC64;

Query Match 44.2%; Score 57; DB 16; Length 412;  
Best Local Similarity 52.2%; Pred. No. 9.7;  
Matches 12; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

OY 4 LEALADQTDALQSEEAAYVKADN 26  
::|||  
DB 49 IQALQEQIDALSSQEKVSKMDN 71

## RESULT 9

ID 0935B1 PRELIMINARY; PRT; 850 AA.  
AC 0935B1;  
DT 01-DEC-2001 (TREMBlrel. 19, Created)  
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE PUTATIVE PHAGE TAIL PROTEIN.  
GN HCM2.0051C.  
OS Salmonella enterica subsp. enterica serovar Typh1.  
OG Plasmid pHCM2.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Salmonella.  
OX NCBI\_TaxID=90370;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-C718;  
RX MEDLINE=21534947; PubMed=11677608;  
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,

RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebalthia M., Baker S., Basham D., Brooks K., Chillingworth T., Connor P., Cronin A., Davies P., Davies R.M., Dowd L., White N., Farrar J., Fellwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagals K., Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C., Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;

RT "Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar Typhimurium CT18.";

RI Nature 413:848-852(2001).;

DR EMBL: AL513384; CAD09918.1; -.

KW Plasmid.

SO SEQUENCE 850 AA; 90917 MW; 21271493A07A6563 CRC64;

Query Match 44.2%; Score 57; DB 2; Length 850;  
Best Local Similarity 59.1%; Pred. No. 21;  
Matches 13; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

OY 6 ALADOTDALQSEEAAYVKADNA 27  
DB 212 AASQTAALQSEQVAAAHADSA 233

RESULT 10  
O931X0 PRELIMINARY; PRT; 999 AA.

AC 0931X0;  
DT 01-DEC-2001 (TREMblrel. 19, Created)  
DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)  
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
DE PUTATIVE EXONUCLEASE.  
GN SCBAC36F5.11C.  
OS Streptomyces coelicolor.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
OX NCBI\_TaxID=1902;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-A3(2);  
RA Saunderson D., Harris D.;  
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-A3(2);  
RA Cereno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;  
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-A3(2);  
RA MEDLINE-97000351; PubMed-8843436;  
RA Redenbach M., Kleier H.M., Denapalte D., Eichner A., Cullum J., Kinashi H., Hopwood D.A.;  
RT "A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb Streptomyces coelicolor A3(2) chromosome.";  
RI Mol. Microbiol. 21:77-96(1996).  
DR EMBL: AL592292; CAC42847.1; -.  
KW Exonuclease.  
SO SEQUENCE 999 AA; 108134 MW; 275A16F51A85C22 CRC64;

Query Match 39.9%; Score 51.5; DB 2; Length 999;  
Best Local Similarity 42.5%; Pred. No. 14e+02;  
Matches 17; Conservative 3; Mismatches 7; Indels 13; Gaps 2;

OY 1 SDALADOT-----DALQSEEAAY---VKADNA 27  
DB 716 ADAALALDDTAHRELQRLDAMQSDDAVRAVLAEDTA 755

RESULT 11  
O92FR3 PRELIMINARY; PRT; 203 AA.  
AC 092FR3;  
DT 01-DEC-2001 (TREMblrel. 19, Created)

DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)  
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
DE LIND040 PROTEIN.  
GN LIND040.  
OS Listeria innocua.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
OC Bacillus/Staphylococcus group; Listeria.  
OX NCBI\_TaxID=1642;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-CLIP 11262 / SEROVAR 6A;  
RX PubMed-11679669;  
RA Glaser P., Frangoul L., Buchrieser C., Rusniok C., Amend A., Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T., Charbit A., Chetoui F., Couve E., de Darvar A., Dehoux P., Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussanget O., Eutlian K.-D., Fehli H., Gomez-Lopez N., Hain T., Haut J., Jackson D., Jones L.-M., Kaerst U., Kretz J., Kuhn M., Kunst F., Kurupkat G., Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H., Nordstiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R., Remmel B., Rose M., Schueter T., Simoes N., Tierrez A., Vazquez-Boland J.-A., Voss H., Weiland J., Cossart P.;

RT "Comparative genomics of *Listeria* species.";  
RI Science 294:849-852(2001).  
DR EMBL: AL596163; CAC95273.1; -.  
DR Listlist: LIND0040; -.  
KW Complete proteome.  
SO SEQUENCE 203 AA; 22611 MW; 0F53A6AC1FC1B0B CRC64;

Query Match 39.5%; Score 51; DB 16; Length 203;  
Best Local Similarity 38.5%; Pred. No. 29;  
Matches 10; Conservative 6; Mismatches 10; Indels 0; Gaps 0;

OY 1 SDALADOTDALQSEEAAYVKADN 26  
DB 109 AETKEADDETDLKEDTGVEKEN 134

RESULT 12  
O86553 PRELIMINARY; PRT; 276 AA.

AC 086553;  
DT 01-NOV-1998 (TREMblrel. 08, Created)  
DT 01-NOV-1998 (TREMblrel. 08, Last sequence update)  
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
DE PUTATIVE DEHYDROGENASE.  
GN SCIF2.16C.  
OS Streptomyces coelicolor.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
OX NCBI\_TaxID=1902;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-A3(2);  
RA Seeger K.J., Harris D.;  
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-A3(2);  
RA Parkhill J., Barrell B.G., Rajandream M.A.;  
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-A3(2);  
RA MEDLINE-97000351; PubMed-8843436;  
RA Redenbach M., Kleier H.M., Denapalte D., Eichner A., Cullum J., Kinashi H., Hopwood D.A.;  
RT "A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb Streptomyces coelicolor A3(2) chromosome.";  
RI Mol. Microbiol. 21:77-96(1996).  
CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES (SDR) FAMILY.

```

DR EMBL: AL031350; CAA20507.1;
DR HSP: 064105; 10AA;
DR InterPro: IPR002198; ADH_short.
DR Pfam: PF00106; adh_short.1.
DR PRINTS: PR00080; SDRFAMILY.
DR PROSITE: PS00061; ADH_SHORT; UNKNOWN_1.
DR Oxidoreductase.
SQ SEQUENCE 276 AA; 29010 MW; 30F003B870D63C7C CRC64;

Query Match 39.5%; Score 51; DB 2; Length 276;
Best Local Similarity 48.0%; Pred. No. 40;
Matches 12; Conservative 3; Mismatches 10; Indels 0; Gaps 0;

QY 1 SDALALADDTALQSEAAVYKAD 25
Db 40 ADALHALDEIRATHSVYVDVIAD 64
:|||||:|:|:|

RESULT 13
O9HIV3 PRELIMINARY; PRT; 486 AA.
ID O9HIV3 PRELIMINARY; PRT; 486 AA.
AC O9HIV3;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-OCT-2001 (TReMBLrel. 18, Last annotation update)
DE L-2,4-DIAMINOBUTYRATE DECARBOXYLASE.
GN BDB OR VNG6211G.
OS Halobacterium sp. (strain NRC-1).
OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;
OC Plasmid pNRC200.
OC Halobacterium.
OX NCBI_TaxID=64091;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20504483; PubMed=11016950;
RA Ng W.V., Kennedy S.P., Mahairas G.G., Bergquist B., Pan M.,
RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Shroana J.,
RA Swartzell S., Welt D., Hall J., Dahl T.A., Weltl R., Goo Y.A.,
RA Leitauer B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angewine C.M., Dale H.,
RA Isenbarger T.A., Peck R.F., Pohlschoder M., Spudis J.L., Jung K.-H.,
RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA Ehardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarina S.,
RA "Genome sequence of Halobacterium species NRC-1."
RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
CC - COFACTOR: PYRIDOXAL PHOSPHATE (BY SIMILARITY).
CC - SIMILARITY: BELONGS TO GROUP II DECARBOXYLASES (DDC, GAD, HDC AND TYRDC).
DR EMBL: AE005155; AAG20873.1;
DR InterPro: IPR002129; Pyridoxal_dec.
DR Pfam: PF00282; Pyridoxal_dec.1.
DR PRINTS: PR00800; YHDCRBOXLASE.
KM Complete proteome; Decarboxylase; Lyase; Plasmid; Pyridoxal phosphate.
SQ SEQUENCE 486 AA; 51092 MW; 9CB8996078AF9A16 CRC64;

Query Match 39.5%; Score 51; DB 17; Length 486;
Best Local Similarity 46.2%; Pred. No. 73;
Matches 12; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

QY 2 DATEALADDTALQSEAAVYKADNA 27
Db 355 DGLGALYDRTIALADVAAGLIRADPA 380
:|||||:|:|:|

RESULT 14
O76506 PRELIMINARY; PRT; 4589 AA.
ID O76506 PRELIMINARY; PRT; 4589 AA.
AC O76506;
DT 01-NOV-1998 (TReMBLrel. 08, Created)
DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE CLINARY OUTER ARM DYNEIN BETA HEAVY CHAIN.
GN DYH4.

OS Tetrahymena thermophila.
OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;
OC Tetrahymenina; Tetrahymena.
OX NCBI_TaxID=5911;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99169020; PubMed=10069817;
RA Lee S., Wisniewski J.C., Dentler W.L., Asai D.J.;
RT "Gene knockouts reveal separate functions for two cytoplasmic dyneins in Tetrahymena thermophila."
RL Mol. Biol. Cell 10:771-784(1999).
RN [2]
RP SEQUENCE FROM N.A.
RA Lincoln L.M., Gibson T.M., Asai D.J., Forney J.D.;
RT "A gene knockout reveals that dynein beta heavy chain is required in Tetrahymena thermophila."
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF072878; AAC26117.1;
DR InterPro: IPR004273; Dynein_heavy.
DR Pfam: PF03028; Dynein_heavy.1.
SQ SEQUENCE 4589 AA; 528391 MW; 7164E57C28A12BC1 CRC64;

Query Match 39.5%; Score 51; DB 5; Length 4589;
Best Local Similarity 60.0%; Pred. No. 8,6e+02;
Matches 12; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 2 DATEALADDTALQSEAAV 21
Db 3205 EATEALAEALPALRSREAAV 3224
:|||||:|:|:|

RESULT 15
Q20439 PRELIMINARY; PRT; 1968 AA.
ID Q20439 PRELIMINARY; PRT; 1968 AA.
AC Q20439; Q22545;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 08, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE MYO-1 PROTEIN.
GN MYO-1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA McMurray A.;
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Alnsough R., Anderson K., Baynes C., Barks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopira A., Saunders D., Showkeen R.,
RA Smalton N., Smith A., Sonnenhammer E., Staden R., Sulston J.,
RA Telford-Mleg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Spratt J., Wohlman P.,
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans."
RL Nature 368:32-38(1994).
DR EMBL: Z68119; CAA92197.1;
DR EMBL: Z68117; CAA92197.1; JOINED.
DR EMBL: Z68117; CAA92183.1;
DR EMBL: Z68119; CAA92183.1; JOINED.
DR HSP: P10587; IBR2.
DR InterPro: IPR001609; myosin_head.
DR InterPro: IPR004009; Myosin_N.
DR InterPro: IPR002928; Myosin_tail.
DR InterPro: IPR000209; Peptidase_S8.

```

```

DR Pfam: PF00063; myosin-head; 1.
DR Pfam: PF02736; myosin_N; 1.
DR Pfam: PF01576; myosin_tail; 2.
DR PRINTS: PR00193; MYOSINHEAVY.
DR ProDom: PD000355; myosin-head; 1.
DR SMART: SM00242; MSC; 1.
DR PROSITE: PS00136; SUBTILASE-ASP; UNKNOWN_1.
SQ SEQUENCE 1968 AA; 225431 MW; 5AA0481693CB6267 CRC64;

Query Match
Best Local Similarity 38.8%; Score 50; DB 5; Length 1968;
Matches 9; Conservative 7; Mismatches 7; Indels 0; Gaps 0;

QY 1 SDALADQTDALQSEEAAYK 23
DB 1213 SDALQSDIQIEQLQKQRIEK 1235

RESULT 16
087677
ID 087677 PRELIMINARY; PRT; 396 AA.
AC 087677:
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE GLYCOSYLTRANSFERASE.
GN BGTEA.
OS Amycolatopsis mediterranei (Nocardia mediterranei).
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Pseudonocardiales; Pseudonocardaceae; Amycolatops.
OX NCBI_TaxID=33910;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM5908;
RX MEDLINE=99318579; PubMed=10390204;
RA Pelzer S., Suesmuth R., Heckmann D., Recktenwald J., Huber P.,
RA Jung G., Wohlleben W.;
RT "Identification and analysis of the balhimycin biosynthetic gene
RT cluster and its use for manipulating glycopeptide biosynthesis in
RT Amycolatopsis mediterranei DSM5908."
RL Antimicrob. Agents Chemother. 43:1565-1573(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM5908;
RA Recktenwald J., Shwaky R.M., Puk O., Pfennig F., Keller U.,
RA Wohlleben W., Pelzer S.;
RT "The nonribosomal biosynthesis of Vancomycin-type antibiotics: A
RT heptapeptide backbone and eight peptide synthetase modules."
RL Submitted (Aug-2001) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM5908;
RA Pfeiffer V., Nicholson G.J., Ries J., Recktenwald J., Schefer A.B.,
RA Shwaky R.M., Schroeder J., Wohlleben W., Pelzer S.;
RT "A Polyketide Synthase of Glycopeptide Biosynthesis: the Biosynthesis
RT of the non-ribosomal Amino Acid (S)-3,5-Dihydroxyphenylglycine."
RL Submitted (Aug-2001) to the EMBL/Genbank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM5908;
RA Puk O., Huber P., Bischoff D., Recktenwald J., Jung G.,
RA Suesmuth R.D., Van Pee K.H., Wohlleben W., Pelzer S.;
RT "Glycopeptide biosynthesis in Amycolatopsis mediterranei DSM5908:
RT function of a halogenase and a haloperoxidase."
RL Submitted (Aug-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL: Y16952; CAA76551.1;
SQ SEQUENCE 396 AA; 41131 MW; 1C40BDEDDDF2B72 CRC64;

Query Match
Best Local Similarity 38.4%; Score 49.5; DB 2; Length 396;
Matches 14; Conservative 3; Mismatches 8; Indels 1; Gaps 1;

```

```

QY 1 SDAL-ALADQTDALQSEEAAYKAD 25
DB 349 SDALFALALPERRARAEVAGTVPRD 374

RESULT 17
09JP85
ID 09JP85 PRELIMINARY; PRT; 207 AA.
AC 09JP85:
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE PUTATIVE TRANSCRIPTIONAL REGULATOR PDHR.
GN PDHR.
OS Streptomyces scouleriensis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycetales; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=73044;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99078078; PubMed=9858775;
RA Youn H., Kwak J., Youn H.D., Hah Y.C., Kang S.O.;
RT "Lipoamide dehydrogenase from Streptomyces scouleriensis: biochemical and
RT genetic properties."
RL Biochim. Biophys. Acta 1388:405-418(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA Youn H., Kang S.O.;
RL Submitted (Apr-1999) to the EMBL/Genbank/DBJ databases.
CC -1 SIMILARITY: BELONGS TO THE GNTR FAMILY OF TRANSCRIPTIONAL
CC REGULATORS.
DR EMBL: AF047034; AAF37157.1;
DR InterPro: IPR001525; C5_DNA_meth.
DR InterPro: IPR000524; HTH_Gntr.
DR Pfam: PF00392; gntr; 1.
DR SMART: SM00345; HTH_GNTR; 1.
DR PROSITE: PS00095; C5_MTAASE_2; UNKNOWN_1.
KW DNA-binding; Transcription regulation.
SQ SEQUENCE 207 AA; 22985 MW; 478003BCAD8B38A4 CRC64;

Query Match
Best Local Similarity 38.0%; Score 49; DB 2; Length 207;
Matches 11; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

QY 3 ALEALADQTDALQSEEAAYKADNA 27
DB 97 AALELAESCEALPHEVAALYEAQDA 121

RESULT 18
092PL0
ID 092PL0 PRELIMINARY; PRT; 251 AA.
AC 092PL0:
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE PUTATIVE ABC TRANSPORTER ATP-BINDING PROTEIN.
GN SMC00531.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21368234; PubMed=11474104;
RA Gallibert F., Finan T.M., Long S.R., Puehler A., Abola P., Ampe F.,
RA Barloy-Hubler F., Barnett M.J., Becker A., Bolstad P., Botte G.,
RA Boutry M., Bowser L., Buhrmester J., Cadieu E., Capela D., Chain P.,
RA Cowie A., Davis R.W., Dreano S., Federspiel N.A., Fisher R.F.,
RA Gloux S., Godrie T., Goffeau A., Golding B., Guzy J., Gurjal M.,
RA Hernandez-Lucas I., Hong A., Huizar L., Hyman R.W., Jones T., Kahn D.,
RA Kahn M.L., Kaiman S., Keating D.H., Kiss E., Komp C., Lelaure V.,

```



RA Masny D., Palm C., Peck M.C., Pohl T.M., Portetelle D., Purnelle B.,  
 RA Ramberger U., Surzycki R., Thebaud P., Vandenol M.,  
 RA Vorreiter F.J., Weidner S., Wells D.H., Wong K.-C., Batut J.,  
 RT "The composite genome of the legume symbiont *Shorhizobium meliloti*."  
 RL Science 293:668-672(2001).  
 DR EMBL: AL591788; CAC46313.1; -.  
 KW Complete proteome.  
 SQ SEQUENCE 251 AA; 27570 MW; A82EFC3007A5FD6 CRC64;

Query Match 38.0%; Score 49; DB 16; Length 251;  
 Best Local Similarity 52.4%; Pred. No. 67;  
 Matches 11; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

OY 2 DALEALADOTDALQSEAAVY 22  
 DB 181 DALKIVADGVNLRSPRAYV 201

RESULT 19  
 O9SVK1 PRELIMINARY; PRT; 445 AA.

AC O9SVK1; PRELIMINARY; PRT; 445 AA.  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)  
 DE HYPOTHETICAL 50.1 KDA PROTEIN.  
 GN F19H22.10 OR AT4G38910.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 NC NCBL\_TaxID=3702;

RN [1]  
 RA SEQUENCE FROM N.A.  
 RA Bevan M., Murphy G., Ridley P., Hudson S., Bancroft I., Mewes H.W.,  
 RA Mayer K.F.X., Schueller C.;  
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RA SEQUENCE FROM N.A.  
 RA EU Arabidopsis sequencing project;  
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RA SEQUENCE FROM N.A.  
 RA Murphy G., Ridley P., Hudson S., Mewes H.W., Lemcke K., Mayer K.F.X.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RA SEQUENCE FROM N.A.  
 RA EU Arabidopsis sequencing project;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AL035679; CAB38811.1; -.  
 DR EMBL: AL161594; CAB80554.1; -.  
 KW Hypothetical protein.  
 SQ SEQUENCE 445 AA; 50125 MW; E70B1F0962985116 CRC64;

Query Match 38.0%; Score 49; DB 10; Length 445;  
 Best Local Similarity 43.5%; Pred. No. 1.3e+02;  
 Matches 10; Conservative 6; Mismatches 7; Indels 0; Gaps 0;

OY 5 EALADQTDALQSEAAVVKADNA 27  
 DB 78 EALASRDEALQGRDKALSERDNA 100

RESULT 20  
 O9HPR2 PRELIMINARY; PRT; 684 AA.

AC O9HPR2; PRELIMINARY; PRT; 684 AA.  
 DT 01-MAR-2001 (TREMBlrel. 16, Created)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE VNG1511C.  
 GN Halobacterium sp. (strain NRC-1).

OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;  
 OC Halobacterium.  
 NC NCBL\_TaxID=64091;  
 RN [1]  
 RA SEQUENCE FROM N.A.  
 RX MEDLINE-20504483; PubMed-11016950;  
 RA M.W.V., Kennedy S.P., Mahairas G.G., Bergquist B., Pan M.,  
 RA Shukla H.D., Lasky S.R., Balliga N.S., Thorsson V., Shrogha J.,  
 RA Swartzell S., Weir D., Hall J., Dahl T.A., Weir R., Goo Y.A.,  
 RA Leitner B., Keller K., Cruz R., Danson M.J., Hough D.W.,  
 RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,  
 RA Isenberger T.A., Peck R.F., Pohlshocher M., Spudich J.L., Jung K.-H.,  
 RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,  
 RA Ehardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.,  
 RT "Genome sequence of Halobacterium species NRC-1."  
 RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).  
 DR EMBL: AE005065; AAG19805.1; -.  
 DR HSSP: P05055; ISRO.  
 DR InterPro: IPR001667; DHH.  
 DR InterPro: IPR003029; SI.  
 DR InterPro: IPR004365; tRNA\_anti.  
 DR Pfam: PF01368; DHH; 1.  
 DR Pfam: PF00575; SI; 1.  
 DR Pfam: PF01336; tRNA\_anti; 1.  
 DR SMART: SM00316; SI; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 684 AA; 73209 MW; 830AB7964BA08713 CRC64;

Query Match 38.0%; Score 49; DB 17; Length 684;  
 Best Local Similarity 46.4%; Pred. No. 2e+02;  
 Matches 13; Conservative 4; Mismatches 9; Indels 2; Gaps 1;

OY 2 DALEALADOTDALQSEAAVVKADNA 27  
 DB 604 DALHRRADSTVDLGGAGEDALRVSDA 631

RESULT 21  
 O94MN6 PRELIMINARY; PRT; 207 AA.

AC O94MN6; PRELIMINARY; PRT; 207 AA.  
 DT 01-DEC-2001 (TREMBlrel. 19, Created)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE P83.  
 OS Bacteriophage Mx8.  
 OC Viruses.  
 NC NCBL\_TaxID=49964;  
 RN [1]  
 RA SEQUENCE FROM N.A.  
 RA Youderian P., Walters D., Salmi D., Magrini V., Hartzell P.L.;  
 RT "Genome Organization of temperate Myxococcus phage Mx8."  
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF396866; AAK94418.1; -.  
 SQ SEQUENCE 207 AA; 21771 MW; 45B1B5B396C448875 CRC64;

Query Match 37.2%; Score 48; DB 9; Length 207;  
 Best Local Similarity 38.2%; Pred. No. 75;  
 Matches 13; Conservative 6; Mismatches 5; Indels 10; Gaps 1;

OY 5 EALADQTD-----ALQSEAAVVKADNA 28  
 DB 48 EALAEHAKAEVAVYARGLAVAEASAAEDNNA 81

RESULT 22  
 O9FHF8 PRELIMINARY; PRT; 1160 AA.

AC O9FHF8; PRELIMINARY; PRT; 1160 AA.  
 DT 01-MAR-2001 (TREMBlrel. 16, Created)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE DISEASE RESISTANCE PROTEIN-LIKE.

```

00 Arabidopsis thaliana (Mouse-ear-cress).
0C Eukaryota: Viridiplantae: Streptophyta; Embryophyta; Tracheophyta;
0C Spermatophyta; Magnoliophyta: eudicotyledons; core eudicots; Rosidae;
0C eurosid II; Brassicales; Brassicaceae; Arabidopsis.
0X NCBI_TaxID=3702;
01 [1]
0P SEQUENCE FROM N.A.
0R STRAIN=COLUMBIA;
0C MEDLINE=20181125;
0A Sato S., Nakamura Y., Kaneko T., Katoh T., Asamizu E., Kotani H.,
0A Tabata S.;
0T RT
0T "Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence
0T features of the regions of 3,076,755 bp covered by sixty pl and TAC
0T clones.";
0L DNA Res. 7:31-63(2000).
0R EMBL: AB019223; BAB10815.1; -.
0R InterPro: IPR003593; AAA.
0R InterPro: IPR001687; ATP-GTP_A.
0R InterPro: IPR007657; Disease_resist.
0R InterPro: IPR001611; LRR.
0R InterPro: IPR002182; NB-ARC.
0R InterPro: IPR000157; TIR.
0R Pfam: PF00560; LRR_4.
0R Pfam: PF00931; NB-ARC; 1.
0R Pfam: PF01582; TIR; 1.
0R PRINTS: PRO0364; DISEASERIST.
0R SMART: SM00382; AAA; 1.
0R SMART: SM00255; TIR; 1.
0K ATP-binding.
00 SEQUENCE 1160 AA; 131691 MW; E287BEA1B19CECA CRC64;

```

Query Match 37.2% Score 48; DB 10; length 1160;  
Best Local Similarity 44.0%; Pred. No. 4.9e+02;  
Matches 11; Conservative 3; Mismatches 11; Indels 0; Gaps 0.

QY 2 DALEALADOTDALOSEAAVVKADN 26  
| | | | | | | | | | : |  
Db 309 DYLDALADOTWFGSGSRITVTEN 333

RESULT	23			
084403				
ID	084403	PRELIMINARY;	PRT;	254 AA.
AC	084403;			
DT	01-NOV-1998 (TEMBREL, 08, Created)			
DT	01-NOV-1998 (TEMBREL, 08, Last sequence update)			
DT	01-DEC-2001 (TEMBREL, 19, Last annotation update)			
DE	HYPOHETICAL 29.6 KDA PROTEIN.			
GN	CT398.			
OS	Chlamydia trachomatis.			
OC	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.			
OX	NCBI_TaxId=813;			
RN	1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=D/UW-3/CX;			
RX	MEDLINE=9900809; PubMed=9784136;			
RA	Stephens R.S., Kaiman S., Lammel C.J., Fan J., Marathe R., Aravind L.,			
RA	Matliger W.P., Ollinger L., Taturov R.L., Zhao Q., Koonin E.V.,			
RA	Davis R.W.;			
RT	"Genome sequence of an obligate intracellular pathogen of humans:			
RT	Chlamydia trachomatis."			
RL	Science 282:754-759(1998).			
DR	EMBL: AE001313; AAC6795.1;			
DR	Interpro: IPR003743; DUF164.			
DR	Pfam: PF02591; DUF164; 1			
KM	Hypothetical protein; Complete proteome.			
QO	SEQUENCE 254 AA; 29581 MW; 49F1BAFCC0651BF CRC64;			

Query Match	36.48;	Score 47;	DB 16;	Length 254;
Best Local Similarity	36.48;	Pred. No. 1.3e+02;		
Matches	8;	Conservative	9;	Mismatches 5;
				Indels 0;
				Gaps 0;
0y	4	LEALADQTDALQSEEAIVKAD	25	

```
Db 67 IQEISDQINKLENOQAAYKKMD 88
```

RESULT 24

ID	PRELIMINARY;	PRT;	254 AA
Q9PJ29			

DT 01-OCT-2000 (Tremblay, 15, Created)

DT 01-DEC-2001 (Tremblay, 19, Last annotation update)

GN TC0677.

OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia

RN	[1]
1	1
2	2
3	3
4	4
5	5
6	6
7	7
8	8
9	9
10	10
11	11
12	12
13	13
14	14
15	15
16	16
17	17
18	18
19	19
20	20
21	21
22	22
23	23
24	24
25	25
26	26
27	27
28	28
29	29
30	30
31	31
32	32
33	33
34	34
35	35
36	36
37	37
38	38
39	39
40	40
41	41
42	42
43	43
44	44
45	45
46	46
47	47
48	48
49	49
50	50
51	51
52	52
53	53
54	54
55	55
56	56
57	57
58	58
59	59
60	60
61	61
62	62
63	63
64	64
65	65
66	66
67	67
68	68
69	69
70	70
71	71
72	72
73	73
74	74
75	75
76	76
77	77
78	78
79	79
80	80
81	81
82	82
83	83
84	84
85	85
86	86
87	87
88	88
89	89
90	90
91	91
92	92
93	93
94	94
95	95
96	96
97	97
98	98
99	99
100	100

RC STRAIN=MOPN / NIGG;

RA Read T.D., Brunham R.C., Shen C.,

RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,

RA Eisen J., Fraser C.M.;

RT pneumoniae AR39, "

DR EMBL; AE002336; AAF39497.1; -.  
CD EICD; EC0677.

```
DR      InterPro; IPR000345; Cytc_heme_bind.  
DR      InterPro; IPR003742; hemeA  
DR      InterPro; IPR003743; hemeA
```

DR Píam; PF02591; DUF164; 1.  
DR PROCTE; PC00100; CYTOCHROME C; UNKNOWN 1

KW	Hypothetical protein; Complete proteome.
CO	SEQUENCE JEA 11. 30636 MW: 80360PDE5700

Queen: March 30, 1880. Eggs: 17. Pup: 15. Length:

Query Match	36.4%	Score 47	DB 16	Length 254
Best Local Similarity	36.4%	Pred.No. 1.3e+02		
Matches	8	Conservative	5	Indels 0
				Gaps 0

QY 4 LEALADQTDALQSEEA VVKAD 25

Db 67 IQEISDQINKLENQQA VKKMD 88

## RESULT 25

Q9KYM4 PRELIMINARY; PRT; 387 AA

DT 01-OCT-2000 (Tremblay, 15, Created)

DT 01-OCT-2000 (Tremblérel. 15, Last annotation update)

DE FORGATTO DECEASED; MARRIED;  
GN SCE33.19.

Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
Bacteroidetes; Actinobacteria; Actinobacteridae;

NCBI\_TaxID=1902; Streptomyces

RP SEQUENCE FROM N.A.

SAUNDERS D.C., HARRIS D.;

AN  
BIBLIOGRAPHY (1981-2000) OF THE FISH/CHANDLER/DEBB  
RN [2]

```
RC STRAIN=A3(2);
```

Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.

SEQUENCE FROM N.A

AC SIMON-HS(2),  
 RX MEDLINE=97000351; Pubmed=8843436;

```

RA Redenbach M., Kieser H.M., Denapate D., Eichner A., Cullum J.,
RT "Aashi H., Hopwood D.A.;
RT "Aashi H., Hopwood D.A.; a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome."
RT Mol. Microbiol. 21:77-96(1996).
DR EMBL: AL355774; CAB90928.1;
SQ SEQUENCE 387 AA; 40633 MW; F6DE33AB67652A57 CRC64;

Query Match
Best Local Similarity 36.4%; Score 47; DB 2; Length 387;
Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 3 ALEALADQTDALQ-SEEAAYKADNA 17
DB 369 ALEALADQTDALQ-SEEAAYKADNA 383

RESULT 26
Q95YX4 PRELIMINARY; PRT; 1617 AA.
AC Q95YX4;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE POSSIBLE LIVER STAGE ANTIGEN.
GN LML2.85.
OS Leishmania major.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
ON NCBI_TaxID=5664;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-FRIEDLIN;
RA Murphy L., Quail M., Harris D., Rajandream M., Ivens A., Barrell B.,
RA Oliver K.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL390114; CAC59787.1;
SQ SEQUENCE 1617 AA; 174798 MW; D7A5E04C928756E5 CRC64;

Query Match
Best Local Similarity 36.4%; Score 47; DB 5; Length 1617;
Matches 12; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

QY 3 ALEALADQTDALQ-SEEAAYKADNA 27
DB 306 ALEALADQTDALQ-SEEAAYKADNA 330

RESULT 27
Q9HPA2 PRELIMINARY; PRT; 536 AA.
AC Q9HPA2;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE HTR17 TRANSDUCER.
GN HTR17 OR VNC1733G.
OS Halobacterium sp. (strain NRC-1).
OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;
OC Halobacterium.
ON NCBI_TaxID=64091;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=20504483; PubMed=11016950;
RA Ng W.V., Kennedy S.P., Mahaitas G.G., Bergquist B., Pan M.,
RA Shukla H.D., Lasky S.R., Ballig N.S., Thorsson V., Sbrogna J.,
RA Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
RA Leitauer B., Keller K., Cruz R., Danon M.J., Hough D.W.,
RA Leddicks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA Isenberger T.A., Beck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,
RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA Ehrhardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.;
RT "Genome sequence of Halobacterium species NRC-1."
RT Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).

```

```

DR EMBL: AE005078; AAG19968.1;
DR HSSP: P02942; 1Q07.
DR InterPro: IPR004089; Chemotaxis_transducer.
DR InterPro: IPR003660; HAMP.
DR InterPro: IPR004090; Me_chemotaxis.
DR Pfam: PF000015; MCPsignal; 1.
DR PRINTS: PR00260; CHEMTRNSDUCR.
DR SMART: SM00304; HAMP; 1.
DR SMART: SM00283; MA; 1.
DR Complete proteome.
SQ SEQUENCE 536 AA; 56815 MW; 3B3D2FD8701EE04 CRC64;

Query Match
Best Local Similarity 36.0%; Score 46.5; DB 17; Length 536;
Matches 13; Conservative 2; Mismatches 10; Indels 1; Gaps 1;

QY 2 DALEALADQTDALQ-SEEAAYKADNA 27
DB 295 DALEALADQTDALQ-SEEAAYKADNA 319

RESULT 28
Q979W1 PRELIMINARY; PRT; 541 AA.
AC Q979W1;
DT 01-OCT-2001 (TREMBlrel. 18, Created)
DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE PYRUVATE KINASE.
GN TVG1076108.
OS Thermoplasma volcanium.
OC Archaea; Euryarchaeota; Thermoplasmatales; Thermoplasmaceae;
OC Thermoplasma.
ON NCBI_TaxID=50339;
RN [1]
RP SEQUENCE FROM N.A.
RC SPRAY-GSSI / DSM 4299 / JCM 9571;
RX MEDLINE=20570466; PubMed=11121031;
RA Kawashima T., Amano N., Koike H., Makino S.-I., Higuchi S.,
RA Kawashima T., Yamamoto Y., Watanabe K., Yamazaki M., Kanehori K., Kawamoto T.,
RA Ninozaki T., Yamamoto Y., Aramaki H., Makino K., Suzuki M.;
RT "Archaeal adaptation to higher temperatures revealed by genomic
RT sequence of Thermoplasma volcanium."
RT Proc. Natl. Acad. Sci. U.S.A. 97:14257-14262(2000).
RL EMBL: AP000995; BAB60191.1;
DR InterPro: IPR001697; Pyruvate_kinase.
DR Pfam: PF00224; PK; 1.
DR Pfam: PF02887; PK_C; 1.
DR PRINTS: PR01050; PYRUVTKINASE.
DR Prodom: PD001009; pyruvate_kinase; 1.
KW Kinase; Complete proteome.
SQ SEQUENCE 541 AA; 56810 MW; A2F5CB05C39D2FE CRC64;

Query Match
Best Local Similarity 36.0%; Score 46.5; DB 17; Length 541;
Matches 12; Conservative 3; Mismatches 8; Indels 1; Gaps 1;

QY 1 SDALALADQTDALQ-SEEAAYK 23
DB 278 SDALALADQTDALQ-SEEAAYK 301

RESULT 29
Q9K205 PRELIMINARY; PRT; 749 AA.
AC Q9K205;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE GLUTAMATE-AMMONIA-LIGASE ADENYLYLTRANSFERASE (FRAGMENT).
GN GLENE.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteriales;

```



OY 2 DALEADOTDALQSEEAAYVKADN 26  
 DB 72 DAIDDCOVDALRELEAETVOAHN 96

## RESULT 33

OYK4B4 PRELIMINARY; PRT; 220 AA.

AC OYK4B4;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
 DE HYPOTHETICAL 22.9 KDA PROTEIN (FRAGMENT).

OS Streptomyces coelicolor.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
 OX NCBI\_Taxid=1902;

RN [1]  
 RC SEQUENCE FROM N.A.  
 RC STRAIN-A3(2);  
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.

RC STRAIN-A3(2);  
 RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;  
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.

RC MEDLINE-97000351; PubMed-8843436;  
 RA Redenbach M., Kleser H.M., Denapate D., Eichner A., Cullum J.,  
 RA Kinashi H., Hopwood D.A.;  
 RT "A set of ordered cosmids and a detailed genetic and physical map for  
 RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";

RL MOL. Microbiol. 21:77-96(1996).  
 DR EMBL; AL359214; CAB94589.1; -.

KW Hypothetical protein.  
 FT NON\_TER  
 SQ SEQUENCE 220 AA; 22937 MW; 33BE44C192CA72FC CRC64;

Query Match 35.7%; Score 46; DB 2; Length 220;  
 Best Local Similarity 50.0%; Pred. No. 1.5e+02;  
 Matches 12; Conservative 1; Mismatches 11; Indels 0; Gaps 0;

OY 5 EALADOTDALQSEEAAYVKADNAA 28  
 DB 156 EAOKDLEALQRAEDAQAKADQGA 179

RESULT 34  
 OYAD77 PRELIMINARY; PRT; 237 AA.

AC OYAD77;  
 DT 01-JUN-2001 (TREMBLrel. 17, Created)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE HYPOTHETICAL 25.0 KDA PROTEIN.

GN SCK13.25.  
 OS Streptomyces coelicolor.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
 OX NCBI\_Taxid=1902;

RN [1]  
 RC SEQUENCE FROM N.A.  
 RC STRAIN-A3(2);  
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.

RA Seeger K.J., Harris D.;  
 RA Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
 RP STRAIN-A3(2);  
 RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;

RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.

RC STRAIN-A3(2);  
 RX MEDLINE-97000351; PubMed-8843436;  
 RA Redenbach M., Kleser H.M., Denapate D., Eichner A., Cullum J.,  
 RA Kinashi H., Hopwood D.A.;  
 RT "A set of ordered cosmids and a detailed genetic and physical map for  
 RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";  
 RL MOL. Microbiol. 21:77-96(1996).  
 DR EMBL; AL512667; CAC21634.2; -.  
 KW Hypothetical protein.  
 SQ SEQUENCE 237 AA; 24957 MW; 388BC4428FEAA224 CRC64;

Query Match 35.7%; Score 46; DB 2; Length 237;  
 Best Local Similarity 58.8%; Pred. No. 1.6e+02;  
 Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 3 ALEALADOTDALQSEEA 19  
 DB 150 ALQALADDTLSDAERA 166

## RESULT 35

OY2AK1 PRELIMINARY; PRT; 250 AA.

AC OY2AK1;  
 DT 01-MAY-1999 (TREMBLrel. 10, Created)  
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)  
 DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)  
 DE HYPOTHETICAL 26.8 KDA PROTEIN (FRAGMENT).

OS Streptomyces lividans.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
 OX NCBI\_Taxid=1916;

RN [1]  
 RC SEQUENCE FROM N.A.  
 RC STRAIN-TK21;  
 RA Isiegas C., Parro V., Mellado R.P.;  
 RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; Y13601; CAA73927.1; -.

KW Hypothetical protein.  
 FT NON\_TER  
 SQ SEQUENCE 250 AA; 26783 MW; B678E58C3DC4A634 CRC64;

Query Match 35.7%; Score 46; DB 2; Length 250;  
 Best Local Similarity 46.4%; Pred. No. 1.7e+02;  
 Matches 13; Conservative 4; Mismatches 9; Indels 2; Gaps 1;

OY 3 ALEALADOTDALQSE-EAAYVKADNAA 28  
 DB 153 AREALSEOTEASKELEIKSSRADKAA 180

## RESULT 36

OYFGP2 PRELIMINARY; PRT; 299 AA.

AC OYFGP2;  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE 11-BETA-HYDROXYSTEROID DEHYDROGENASE-LIKE.

OS Arabidopsis thaliana (Mouse-ear cross).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC Eucosids; II; Brassicales; Brassicaceae; Arabidopsids.  
 OX NCBI\_Taxid=3702;

RN [1]  
 RC SEQUENCE FROM N.A.  
 RC STRAIN-COLUMBIA;  
 RA Kaneko T., Katoh T., Asamizu E., Sato S., Nakamura Y., Kotani H.,  
 RA Tabata S.;  
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. XI.;"

RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES  
 CC (SDR) FAMILY.  
 DR EMBL: AB025619; BAB09144.1; -  
 DR HSSP: P50162; 1AE1.  
 DR InterPro: IPR002198; ADH\_short.  
 DR Pfam: PF00106; adh\_short; 1.  
 DR PRINTS: PR00080; SDRFAMILY.  
 DR PROSITE: PS00061; ADH\_SHORT; UNKNOWN\_1.  
 DR Oxidoreductase.  
 KW OXIDOREDUCTASE.  
 SO SEQUENCE 299 AA; 33247 MW; 77F6AC14AD01DD9D CRC64;

Query Match 35.7%; Score 46; DB 10; Length 299;  
 Best Local Similarity 41.7%; Pred. No. 2.1e+02;  
 Matches 10; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

OY 2 DALEALADOTDALQSEAAVVKAD 25  
 DB 82 DRLOVVAADRCRKLGSPPVAAYVRGD 105

RESULT 37  
 O9LUF2  
 ID O9LUF2 PRELIMINARY; PRT; 303 AA.  
 AC O9LUF2;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE 11-BETA-HYDROXYSTEROID DEHYDROGENASE-LIKE.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsia.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=COLUMBIA;  
 RX Sato S., Nakamura Y., Kaneko T., Katoh T., Asamizu E., Kotani H.,  
 RA Tabata S.;  
 RT \*Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence  
 RT features of the regions of 3,076,755 bp covered by sixty P1 and TAC  
 RT clones.;  
 RL DNA Res. 7:31-63(2000).  
 CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES  
 CC (SDR) FAMILY.  
 CC EMBL: AB023037; BAA96982.1; -  
 DR HSSP: P50162; 1AE1.  
 DR InterPro: IPR002198; ADH\_short.  
 DR Pfam: PF00106; adh\_short; 1.  
 DR PRINTS: PR00080; SDRFAMILY.  
 DR PROSITE: PS00061; ADH\_SHORT; UNKNOWN\_1.  
 KW Oxidoreductase.  
 SO SEQUENCE 303 AA; 33689 MW; 1BA39BAE1DDAC859 CRC64;

Query Match 35.7%; Score 46; DB 10; Length 303;  
 Best Local Similarity 41.7%; Pred. No. 2.1e+02;  
 Matches 10; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

OY 2 DALEALADOTDALQSEAAVVKAD 25  
 DB 82 DRLOVVAADRCRKLGSPPVAAYVRGD 105

RESULT 38  
 O9IGV4  
 ID O9IGV4 PRELIMINARY; PRT; 341 AA.  
 AC O9IGV4;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE ENVELOPE GLYCOPROTEIN EI (FRAGMENT).  
 DR SagiYama Vltus.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;  
 OC Alphavirus.  
 OX NCBI\_TaxID=59303;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=TAIWAN;  
 RA Wang H.W., Wang Y.M., Lin H.C., Weng M.H., Lien J.C., Shiao M.F.,  
 RA Yao C.W.;  
 RT Isolation and Characterization of SagiYama Virus, a Member of the  
 RT Semliki Forest Virus Complex, in North Taiwan.\*;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF242890; AAF91319.1; -  
 DR InterPro: IPR002548; Alpha\_E1\_glycop.  
 DR Pfam: PF01589; Alpha\_E1\_glycop; 1.  
 FT NON\_TER 1  
 SO SEQUENCE 341 AA; 36532 MW; D29B6E91DF15207 CRC64;

Query Match 35.7%; Score 46; DB 12; Length 341;  
 Best Local Similarity 41.7%; Pred. No. 2.4e+02;  
 Matches 10; Conservative 4; Mismatches 10; Indels 0; Gaps 0;

OY 5 EALADOTDALQSEAAVVKADNA 28  
 DB 8 EAYVRSVDCRKHDAAYKATRA 31

RESULT 39  
 O9ROV2  
 ID O9ROV2 PRELIMINARY; PRT; 345 AA.  
 AC O9ROV2;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE PUTATIVE MANNOSYLTRANSFERASE (GLYCOSYL TRANSFERASE, GROUP 1 FAMILY  
 DE PROTEIN LPSD).  
 GN LPSD OR CCI013.  
 OS Caulobacter crescentus.  
 OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;  
 OC Caulobacter.  
 OX NCBI\_TaxID=69394;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 19089 / CB15, AND NA1000;  
 RX MEDLINE=98292737; PubMed=9620954;  
 RA Awram P., Smit J.;  
 RT \*The Caulobacter crescentus paracrystalline S-layer protein is  
 RT secreted by an ABC transporter (type I) secretion apparatus.\*;  
 RL J. Bacteriol. 180:3062-3069(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 19089 / CB15, AND NA1000;  
 RA Awram P., Smit J.K.;  
 RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 19089 / CB15, AND NA1000;  
 RA Awram P.;  
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 19089 / CB15;  
 RX MEDLINE=11259647;  
 RA Nierman M.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,  
 RA Elesen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,  
 RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,  
 RA Deboy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,  
 RA Kolony J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,  
 RA Kuterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,  
 RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;  
 RT \*Complete genome sequence of Caulobacter crescentus.\*;  
 RT proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).  
 DR EMBL: AF062345; AAF03166.1; -  
 DR EMBL: AE005779; AAK22297.1; -

DR TIGR; CC1013; Glycosyltransferase; Complete proteome.  
 DR InterPro: IPR001296; Glycosyltransferase; Complete proteome.  
 DR Pfam: PF00534; Glycosyltransferase; Complete proteome.  
 KW Transferrase; Glycosyltransferase; Complete proteome.  
 SQ SEQUENCE 345 AA; 39105 MW; 5E3B5E45D4280449 CRC64;

Query Match 35.7%; Score 46; DB 16; Length 345;  
 Best Local Similarity 43.5%; Pred. No. 2.4e+02;  
 Matches 10; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

OY 6 ALADQTDALQSEEAAYVKKADNAA 28  
 |||:| | | :| | | :  
 Db 303 ALAQFDRLHADKKAATAKMGTA 325

## RESULT 40

O9E338 ID O9E338 PRELIMINARY; PRT; 351 AA.  
 AC O9E338;  
 DT 01-MAR-2001 (T-EMBLrel. 16, Created)  
 DT 01-MAR-2001 (T-EMBLrel. 16, Last sequence update)  
 DT 01-JUN-2001 (T-EMBLrel. 17, Last annotation update)  
 DE POLYPROTEIN.  
 OS Alphavirus M1.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;  
 OC Alphavirus.  
 OX NCBI\_TaxID:97469;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Liang G., Zhao W., Zhou G., He H., Fu S., Li L., Jin Q., Fang M.,  
 RA Hou Y.;  
 RT "Cloning and primary analysis of 3' end of two alphaviruses from  
 RT Hainan Island";  
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AE268025; AAG21223.1;  
 DR InterPro: IPR002548; Alpha\_E1\_glycop.  
 DR Pfam: PF01589; Alpha\_E1\_glycop; 1.  
 SQ SEQUENCE 351 AA; 37679 MW; C1691D77E1E6FDF CRC64;

Query Match 35.7%; Score 46; DB 12; Length 351;  
 Best Local Similarity 41.7%; Pred. No. 2.5e+02;

Matches 10; Conservative 4; Mismatches 10; Indels 0; Gaps 0;

OY 5 EALADQTDALQSEEAAYVKKADNAA 28  
 |||:| | | :| | | :  
 Db 18 EAYVDSQVCKHDAAYKKAHTAA 41

Search completed: October 13, 2002, 04:48:07  
 Job time : 13.1283 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2002, 03:33:22 : Search time 4.04278 Seconds  
(without alignments)  
268.169 Million cell updates/sec

Title: US-09-847-539a-6\_COPY\_59\_86  
Perfect score: 129  
Sequence: 1 SDALEALADQTDALQSEEAAYKADNAA 28

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues  
Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	57.5	44.6	309	1	SANT_PLAFN
2	51	39.5	677	1	YD64_MYCPN
3	50	38.8	617	1	ASMA_ECOLI
4	50	38.8	1947	1	MYSC_CABEL
5	47.5	36.8	421	1	EXOF_RHIME
6	47	36.4	139	1	PAND_MYCTU
7	47	36.4	192	1	NIFE_FRAAL
8	47	36.4	1014	1	MLUB_MYCLE
9	47	36.4	2418	1	SPCA_HUMAN
10	46.5	36.0	544	1	KPYK_THEAC
11	46	35.7	127	1	PAND_STRAM
12	46	35.7	425	1	IM44_CABEL
13	46	35.7	763	1	HTR2_HALNI
14	46	35.7	764	1	HTR2_HALSA
15	46	35.7	787	1	YC88_MYCPN
16	46	35.7	1000	1	Y182_STRCO
17	45	34.9	165	1	MAX_BRARE
18	45	34.9	453	1	KAR1_YEAST
19	45	34.9	474	1	KCC4_RAT
20	45	34.9	476	1	PPBH_PSEAE
21	44.5	34.5	516	1	RSP3_CHLRE
22	44.5	34.5	778	1	HTR6_HALNI
23	44.5	34.5	778	1	HTR6_HALSA
24	44	34.1	251	1	ABCX_ANTSP
25	44	34.1	618	1	CHEA_LISIN
26	44	34.1	618	1	CHEA_LISMO
27	44	34.1	1938	1	MYSD_CABEL
28	43.5	33.7	388	1	MRP4_STRPY
29	43.5	33.7	420	1	YTF3_RHLIT
30	43.5	33.7	471	1	PFB_ECOLI
31	43.5	33.7	1230	1	SMC3_YEEST
32	43	33.3	248	1	RL4_HALHA
33	43	33.3	372	1	TOLA_HAEIN

34	43	33.3	587	1	UL25_HSEVB
35	43	33.3	621	1	YQ01_CABEL
36	43	33.3	627	1	DNM_RICPR
37	43	33.3	697	1	MEPI_LYCER
38	43	33.3	759	1	METE_MYCTU
39	43	33.3	794	1	YB52_MYCPN
40	43	33.3	964	1	MLLS_MYCTU
41	43	33.3	1966	1	MYSB_CABEL
42	43	33.3	2230	1	GOG4_HUMAN
43	43	33.3	2415	1	SPCA_MOUSE
44	43	33.3	6659	1	NEBU_MOUSE
45	42.5	32.9	496	1	NIFE_RHISN

## ALIGNMENTS

## RESULT 1

SANT\_PLAFN STANDARD: PRT: 309 AA.

AC P04928; 13-AUG-1987 (Rel. 05, Created)

DT 13-AUG-1987 (Rel. 05, last sequence update)

DT 01-JUN-1994 (Rel. 29, last annotation update)

DE S-antigen protein precursor.

OS Plasmodium falciparum (isolate NF7 / Ghana).

OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

OX NCBI\_TaxID:5842;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE:85176931; PubMed-3886159;

RA Common A.F., Saint R.B., Coppel R.L., Brown G.V., Anders R.F.,

RT Kemp D.J.;

KT "Conserved sequences flank variable tandem repeats in two S-antigen

genes of Plasmodium falciparum.";

RL Cell 40:775-783(1985).

CC -!- FUNCTION: S ANTIGENS ARE SOLUBLE HEAT-STABLE PROTEINS PRESENT

CC IN THE SERA OF SOME INFECTED INDIVIDUALS.

CC -!- SUBCELLULAR LOCATION: PARASITOPROUS VACUOLE.

CC -!- POLYMORPHISM: DIVERSITY IN S-ANTIGEN IS MAINLY DUE TO POLYMORPHISM

CC IN THE REPEATIVE REGIONS.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL: M10130; AAA29758.1; -.

DR PIR: B22011; YAZON7.

KW Malaria; Repeat; Signal.

FT SIGNAL 1 23

FT CHAIN 24 309

FT DOMAIN 24 56

FT DOMAIN 58 86

FT DOMAIN 97 256

FT DOMAIN 257 286

FT REPEAT 257 271

FT REPEAT 272 286

FT SEQUENCE 309 AA; 33694 MW; 63574C51D62024FC CRC64;

SO

Query Match 44.6% Score 57.5; DB 1; Length 309;

Best Local Similarity 50.0%; Pred. No. 0.72;

Matches 14; Conservative 6; Mismatches 7; Indels 1; Gaps 1;

QY 1 SDALEAL-ADQTDALQSEEAAYKADNAA 27

DB 97 SDALEALSKDEALAKSKDEAKRSDEA 124

```

RESULT 2
YD64_MYCPN          STANDARD:          PRT:   677 AA.
ID   YD64_MYCPN
AC   P75417
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical protein MPN364 (H91_orf677).
GN   MPN364 OR MP472.
CN
OS   Mycoplasma pneumoniae.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC   Mycoplasmales; Mycoplasma.
OX   NCBI_TaxID=2104;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 29342 / M129;
RX   MEDLINE=97105885; PubMed=8948633;
RA   Himmelfeldt R., Hilbert H., Plagens H., Pirkel E., Li B.-C.,
RA   Herrmann R.;
RT   "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT   pneumoniae."
RL   Nucleic Acids Res. 24:4420-4449(1996).
CC   -1- SIMILARITY: BELONGS TO THE MG185 / MG260 FAMILY.
CC   -----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@sib-sib.ch).
CC   -----
DR   EMBL: AE00047; AAB96120.1;
DR   Hypothetical protein: Complete proteome.
KM   SEQUENCE 677 AA: 75591 MW: 87858686079040 CRC64;
SQ
Query Match          39.5%; Score 51; DB 1; Length 677;
Best Local Similarity 50.0%; Pred. No. 13;
Matches 11; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

OY   3 ALEALADOTDALQSEEAAYVKA 24
DB   486 ATEEVLQESNTLQSEEAAYLPA 507

RESULT 3
ASMA_ECOLI          STANDARD:          PRT:   617 AA.
ID   ASMA_ECOLI
AC   P28249; P76390;
DT   01-DEC-1992 (Rel. 24, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Protein asma precursor.
GN   ASMA OR B2084.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE OF 1-379 FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=96065706; PubMed=7476172;
RA   Misra R., Miao Y.;
RT   "Molecular analysis of asma, a locus identified as the suppressor of
RT   OmpF assembly mutants of Escherichia coli K-12."
RL   Mol. Microbiol. 16:779-788(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,

```

```

RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12."
RL   Science 277:1453-1474(1997).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=97251358; PubMed=9097040;
RA   Itoh T., Alba H., Baba T., Fujita K., Hayashi K., Inada T.,
RA   Isono K., Kasai H., Kimura S., Kitakawa M., Kitagawa M.,
RA   Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA   Nakade S., Nakamura Y., Nashimoto H., Nishio Y., Oshima T.,
RA   Saito N., Stampel G., Seki Y., Sivasubraman S., Tagami H.,
RA   Takeda J., Takemoto K., Wada C., Yamamoto Y., Horikuchi T.;
RT   "A 460-kb DNA sequence of the Escherichia coli K-12 genome
RT   corresponding to the 40.1-50.0 min region on the linkage map."
RL   DNA Res. 3:379-392(1996).
RN   [4]
RP   SEQUENCE OF 1-200 FROM N.A.
RX   MEDLINE=92380941; PubMed=1324907;
RA   Wang L., Weiss B.;
RT   "dcd (dCtp deaminase) gene of Escherichia coli: mapping, cloning,
RT   sequencing, and identification as a locus of suppressors of lethal
RT   dut (dUtpase) mutations."
RL   J. Bacteriol. 174:5647-5653(1992).
CC   -1- FUNCTION: INVOLVED IN THE INHIBITION OF ASSEMBLY OF MUTANT OMP
CC   PROTEINS. IN GENERAL, COULD BE INVOLVED IN THE ASSEMBLY OF OUTER
CC   MEMBRANE PROTEINS.
CC   -1- SUBCELLULAR LOCATION: Periplasmic (potential).
CC   -----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@sib-sib.ch).
CC   -----
DR   EMBL: U11035; AAD14778.1;
DR   EMBL: AE000296; AAC75125.1;
DR   EMBL: D90844; BAA15917.1;
DR   EMBL: D90845; BAA15922.1;
DR   EMBL: M90069; AAA23670.1;
DR   PIR: B42940; B42940.
DR   EcoGene: EG11361; asma.
KW   Periplasmic; Signal; Complete proteome.
FT   SIGNAL          1..22
FT   CHAIN           23..617
FT   CONFLICT        354..379
FT   FT              354..379
FT   FT              NEFYRCCHANDQGVGFACNVSTAGT (IN REF. 1).
SQ   SEQUENCE 617 AA: 68962 MW: 23AD060D864BC680 CRC64;

Query Match          38.8%; Score 50; DB 1; Length 617;
Best Local Similarity 45.8%; Pred. No. 16;
Matches 11; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

OY   3 ALEALADOTDALQSEEAAYVKA 26
DB   109 AVIQLTPQTEAVRSDDAPVAPRDN 132

RESULT 4
MYSC_CAEEL          STANDARD:          PRT:   1947 AA.
ID   MYSC_CAEEL
AC   P12845;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-OCT-1989 (Rel. 12, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Myosin heavy chain C (MHC C).
GN   MYO-2.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditidae; Rhabditidae;
OC   Rhabditidae; Pelodierinae; Caenorhabditis.

```

OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RX MEDLINE=89178677; PubMed=2926820;  
 RA Dibb N.J., Maruyama I.N., Krause M., Karn J.;  
 RT "Sequence analysis of the complete *Caenorhabditis elegans* myosin heavy chain gene family";  
 RL J. Mol. Biol. 205:603-613(1989).  
 RN [2]  
 RP SEQUENCE OF 1-21; 112-371 AND 1501-1772 FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RX MEDLINE=85201409; PubMed=388374;  
 RA Karn J., Dibb N.J., Miller D.M.;  
 RT "Cloning nematode myosin genes";  
 RL Cell Muscle Motil. 6:185-237(1985).  
 CC -1- FUNCTION: MUSCLE CONTRACTION.  
 CC -1- SUBUNIT: MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2 HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC) AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).  
 CC -1- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.  
 CC -1- TISSUE SPECIFICITY: FOUND EXCLUSIVELY IN THE PHARYNGEAL MUSCLE.  
 CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES, CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.  
 CC -1- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.  
 CC -1- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT MEROMYOSIN (LM) AND 1 HEAVY MEROMYOSIN (HM). IT CAN LATER BE SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED SUBFRAGMENT (S2).  
 CC -1- MISCELLANEOUS: THERE ARE FOUR DIFFERENT MYOSIN HEAVY CHAINS IN *C. ELEGANS*.  
 CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X08066; CAA30855.1; -;  
 DR EMBL: M37233; AAA28121.1; -;  
 DR EMBL: M37235; AAA28122.1; -;  
 DR EMBL: M37236; AAA28123.1; -;  
 DR PIR: S05697; S05697.  
 DR HSSP: P08799; LMN.  
 DR InterPro: IPR004009; Myosin\_N.  
 DR InterPro: IPR002928; Myosin\_Tail.  
 DR InterPro: IPR001609; myosin\_head.  
 DR Pfam: PF00063; myosin\_N\_head.1.  
 DR Pfam: PF02736; Myosin\_N.1.  
 DR Pfam: PF01576; Myosin\_Tail.1.  
 DR PRINTS: PR00193; MYOSINHEAVY.  
 DR ProDom: PD000355; myosin\_head.1.  
 DR SMART: SM00442; Msc.1.  
 DR Myosin: Muscle protein; Coiled coil; Thick filament; Actin-binding;  
 KW ATP-binding; Methylation; Alkylation; Multigene family.  
 FT DOMAIN 1 855 MYOSIN HEAD-LIKE  
 FT DOMAIN 856 1947 RODLIKE TAIL (S2 AND LMN DOMAINS).  
 FT NP\_BIND 856 1947 COILED COIL (POTENTIAL).  
 FT DOMAIN 174 181 ATP (BY SIMILARITY).  
 FT DOMAIN 668 690 ACTIN-BINDING.  
 FT MOD\_RES 773 787 METHYLATION (SH-1) (POTENTIAL).  
 FT MOD\_RES 125 125 ALKYLATION (SH-2).  
 FT MOD\_RES 708 708 ALKYLATION (SH-2).  
 FT MOD\_RES 718 718 ALKYLATION (SH-2).  
 FT CONFLICT 132 132 E -> D (IN REF. 2).  
 FT CONFLICT 137 137 M -> I (IN REF. 2).  
 SQ SEQUENCE 1947 AA; 223009 MW; 6D65A599BD013627 CRC64;

Query Match 38.8%; Score 50; DB 1; Length 1947;  
 Best Local Similarity 39.1%; Pred. No. 53;  
 Matches 9; Conservative 7; Mismatches 7; Indels 0; Gaps 0;  
 QY 1 SDALADQTDALQSEAAVYK 23  
 DB 1213 SDALQELSDQELQKGRLEK 1235  
 RESULT 5  
 EXOF\_RHIME STANDARD; PRT; 421 AA.  
 AC 002728;  
 DT 01-JUL-1993 (Rel. 26, Created)  
 DT 01-JUL-1993 (Rel. 26, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Exopolysaccharide production protein exof precursor.  
 GN EXOF OR RB1068 OR SMB20945.  
 OS Rhizobium meliloti (Sinorhizobium meliloti).  
 OC plasmid pSymb (megaplasmid 2).  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Rhizobiaceae; Sinorhizobium.  
 OX NCBI\_TaxID=382;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-RCR2011 / S047;  
 RX MEDLINE=93177026; PubMed=8439670;  
 RA Mueller P., Keller M., Weng W.M., Quandt J., Arnold W., Puehler A.;  
 RT "Genetic analysis of the Rhizobium meliloti exoYQ operon: ExoY is homologous to sugar transferases and ExoQ represents a transmembrane protein";  
 RL Mol. Plant Microbe Interact. 6:55-65(1993).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-1021;  
 RX MEDLINE=21396508; PubMed=11481431;  
 RA Finan T.M., Weidner S., Wong K., Buhrmester J., Chain P., Vorholster F.J., Hernandez-Lucas I., Becker A., Cowie A., Gouy J., Golding B., Puehler A.;  
 RT "The complete sequence of the 1.683-kb pSymb megaplasmid from the N2-fixing endosymbiont Sinorhizobium meliloti";  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9889-9894(2001).  
 CC -1- FUNCTION: INVOLVED IN SUCCINOGLYCAN (EPS I) SYNTHESIS. NEEDED FOR THE ADDITION OF THE FIRST SUGAR (GALACTOSE) TO THE ISOPRENOID CARRIER.  
 CC -1- PATHWAY: EXOPOLYSACCHARIDE BIOSYNTHESIS.  
 CC -1- SUBCELLULAR LOCATION: Periplasmic (Probable).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: L05588; AAA26265.1; -;  
 DR EMBL: AL603645; CAC49468.1; -;  
 KW Exopolysaccharide synthesis; Plasmid; Periplasmic; signal;  
 KW Complete proteome.  
 FT SIGNAL 1 31 POTENTIAL.  
 FT CHAIN 32 421 EXOPOLYSACCHARIDE PRODUCTION PROTEIN  
 FT FT EXOF.  
 SQ SEQUENCE 421 AA; 45871 MW; 55D8C5B3A469EF CRC64;  
 QY 4 LEALADQTDALQSEAAVYKADNA 28  
 DB 240 LDALADLRSLQSELEALAKKAETQA 265  
 Query Match 36.8%; Score 47.5; DB 1; Length 421;  
 Best Local Similarity 53.8%; Pred. No. 23;  
 Matches 14; Conservative 2; Mismatches 9; Indels 1; Gaps 1;

```

RESULT 6
PAND_MYCTU          STANDARD:          PRT:          139 AA.
AC 006281:
DT 01-NOV-1997 (rel. 35, Created)
DT 01-NOV-1997 (rel. 35, Last sequence update)
DT 16-OCT-2001 (rel. 40, Last annotation update)
DE Aspartate 1-decarboxylase precursor (EC 4.1.1.11) (Aspartate alpha-
DE decarboxylase).
GN PAND OR RV3601C OR MT3706.1 OR MTC0747B.21.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98293987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeyer K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutherford S., Seeger K., Skellon S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrett B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
RN (2)
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains."
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: L-aspartate -> beta-alanine + CO(2).
CC -1- COFACTOR: Pyruvoyl group (by similarity).
CC -1- PATHWAY: Pantothenate biosynthesis; second branch.
CC -1- SIMILARITY: BELONGS TO THE PAND FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb.ch).
CC -----
DR EMBL: 295557; CAB08943.1; -
DR EMBL: AE007170; AKA48064.1; -
DR HSSP: P31664; 1AW0.
DR TIGR: MT3706.1; -
DR Tuberculist; RV3601C; -
DR InterPro: IPR003190; Asp_decarbox.
DR Pfam: PF02261; Asp_decarbox; 1.
KW Pantothenate biosynthesis; Lyase; Decarboxylase; Pyruvate; Zymogen;
KW Complete proteome.
FT CHAIN 1 24 ASPARTATE 1-DECARBOXYLASE BETA CHAIN (BY
FT SIMILARITY).
FT CHAIN 25 139 ASPARTATE 1-DECARBOXYLASE ALPHA CHAIN (BY
FT MOD_RES 25 25 SIMILARITY).
FT SEQUENCE 139 AA; 14885 MW; C5BPDIC996ED9C6 CRC64;
Query Match 36.4%; Score 47; DB 1; Length 139;
Best Local Similarity 40.9%; Pred. No. 8.6;

```

```

Matches 9; Conservative 3; Mismatches 10; Indels 0; Gaps 0;
QY 7 LADQTDALQSEFAAYVKADNAA 28
Db 32 LMDADLLEGEQVTVIDNCA 53
RESULT 7
NIFZ_FRAAL          STANDARD:          PRT:          192 AA.
AC P46040:
DT 01-NOV-1995 (rel. 32, Created)
DT 01-NOV-1995 (rel. 32, Last sequence update)
DT 01-OCT-1996 (rel. 34, Last annotation update)
DE NifZ protein.
GN NIFZ.
OS Frankia alni.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Frankineae; Frankiaceae; Frankia.
OX NCBI_TaxID=1859;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CP11;
RX MEDLINE=95369734; PubMed=7642138;
RA Harriott O.T., Hosted T.J., Benson D.R.;
RT "Sequences of nifH, nifW, nifZ, nifB and two ORF in the Frankia
RT nitrogen fixation gene cluster."
RL Gene 161:63-67(1995).
CC -1- FUNCTION: UNKNOWN.
CC -1- SIMILARITY: BELONGS TO THE NIFZ FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb.ch).
CC -----
DR EMBL: L29299; AAC82974.1; -
KW Nitrogen fixation.
SQ SEQUENCE 192 AA; 19002 MW; F226C72F59F50B7A CRC64;
Query Match 36.4%; Score 47; DB 1; Length 192;
Best Local Similarity 52.0%; Pred. No. 12;
Matches 13; Conservative 1; Mismatches 11; Indels 0; Gaps 0;
QY 4 LEMADQTDALQSEFAAYVKADNAA 28
Db 74 LEMADQTDALQSEFAAYVKADNAA 98
RESULT 8
MMBL_MYCLE          STANDARD:          PRT:          1014 AA.
AC 006079;
DT 30-MAY-2000 (rel. 39, Created)
DT 30-MAY-2000 (rel. 39, Last sequence update)
DT 16-OCT-2001 (rel. 40, Last annotation update)
DE Putative membrane protein mmp11.
GN MPP11 OR ML2617 OR MLC622.16C.
OS Mycobacterium leprae.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TN;
RX MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eiglmeyer K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthey S., Feltwell T., Fraser A., Hamlin N.,

```

RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,  
 RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,  
 RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,  
 RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,  
 RA Barrett B.G.;  
 RT "Massive gene decay in the leprosy bacillus.";  
 RL Nature 409:1007-1011(2001).  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
 CC -1- SIMILARITY: BELONGS TO THE MPM1 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sid.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: Z59398; CAB0803.1; -  
 DR EMBL: AL583926; CAC32149.1; ALT\_INIT.  
 DR Leproma: ML2617; -  
 DR InterPro: IPR000731; HMGR\_patched\_5TM.  
 DR PROSITE: PS0156; SSO; 1.  
 KW Hypothetical protein; Transmembrane; Complete proteome.  
 FT TRANSMEM 13 33 POTENTIAL.  
 FT TRANSMEM 188 208 POTENTIAL.  
 FT TRANSMEM 214 234 POTENTIAL.  
 FT TRANSMEM 235 255 POTENTIAL.  
 FT TRANSMEM 279 299 POTENTIAL.  
 FT TRANSMEM 311 331 POTENTIAL.  
 FT TRANSMEM 373 393 POTENTIAL.  
 FT TRANSMEM 409 429 POTENTIAL.  
 FT TRANSMEM 530 550 POTENTIAL.  
 FT TRANSMEM 560 580 POTENTIAL.  
 FT TRANSMEM 598 618 POTENTIAL.  
 FT TRANSMEM 649 669 POTENTIAL.  
 FT TRANSMEM 671 691 POTENTIAL.  
 SQ SEQUENCE 1014 AA; 109875 MW; 58B0D79AF56D79 CRC64;  
 Query Match 36.4%; Score 47; DB 1; Length 1014;  
 Best Local Similarity 62.5%; Pred. No. 69;  
 Matches 10; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
 QY 7 LADQTDALQSEEAAYV 22  
 Db 932 LADMDAIEPETAAYV 947  
 SPCA\_HUMAN STANDARD; PRT; 2418 AA.  
 ID SPCA\_HUMAN STANDARD; PRT; 2418 AA.  
 AC P02549; Q15514;  
 DT 21-JUL-1996 (Rel. 01, Created)  
 DT 01-NOV-1996 (Rel. 32, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Spectrin alpha chain, erythrocyte (Erythroid alpha-spectrin).  
 GN SPTA1 OR SPTA.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxId=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-90170949; PubMed-1689726; Scarpa A.L., Coupal E., Leto T.L.,  
 RA Sahr K.E., Laurila P., Kotula L., Winkelmann J.C., Spelcher D.W., Marchesi V.T.,  
 RA Linenbach A.J., Winkelman J.C., Spelcher D.W., Marchesi V.T.,  
 RA Curtis P.J., Forget B.G.;  
 RT "The complete cDNA and polypeptide sequences of human erythroid  
 alpha-spectrin.";  
 RL J. Biol. Chem. 265:4434-4443(1990).  
 RN [2]  
 RP SEQUENCE OF 7-533 FROM N.A., AND VARIANTS PRO-260; PRO-261 AND  
 RP PRO-471.

RX MEDLINE-90093318; PubMed-2794061;  
 RA Sahr K.E., Tobe T., Scarpa A., Laughinghouse K., Marchesi S.L.,  
 RA Agre P., Linenbach A.J., Marchesi V.T., Forget B.G.;  
 RT "Sequence and exon-intron organization of the DNA encoding the alpha  
 RT I domain of human spectrin. Application to the study of mutations  
 RT causing hereditary elliptocytosis.";  
 RL J. Clin. Invest. 84:1243-1252(1989).  
 RN [3]  
 RP SEQUENCE OF 7-601.  
 RX MEDLINE-84087888; PubMed-6654896;  
 RA Spelcher D.W., Davis G., Marchesi V.T.;  
 RT "Structure of human erythrocyte spectrin. II. The sequence of the  
 RT alpha-I domain.";  
 RL J. Biol. Chem. 258:14938-14947(1983).  
 RN [4]  
 RP SEQUENCE OF 7-125.  
 RX MEDLINE-84087887; PubMed-6654895;  
 RA Spelcher D.W., Davis G., Yurchenco P.D., Marchesi V.T.;  
 RT "Structure of human erythrocyte spectrin. I. Isolation of the alpha-I  
 RT domain and its cyanogen bromide peptides.";  
 RL J. Biol. Chem. 258:14931-14937(1983).  
 RN [5]  
 RP SEQUENCE OF 320-450 FROM N.A.  
 RX MEDLINE-86205962; PubMed-3458204;  
 RA Linenbach A.J., Spelcher D.W., Marchesi V.T., Forget B.G.;  
 RT "Cloning of a portion of the chromosomal gene for human erythrocyte  
 RT alpha-spectrin by using a synthetic gene fragment.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 83:2397-2401(1986).  
 RN [6]  
 RP PARTIAL SEQUENCE.  
 RX MEDLINE-84295638; PubMed-6472478;  
 RA Spelcher D.W., Marchesi V.T.;  
 RT "Erythrocyte spectrin is comprised of many homologous triple helical  
 RT segments.";  
 RL Nature 311:177-180(1984).  
 RN [7]  
 RP IDENTIFICATION OF PROBABLE FRAMESHIFT IN 2407-2418.  
 RA Gibson T.J.;  
 RL Unpublished observations (MAR-1995).  
 RN [8]  
 RP REVIEW ON VARIANTS.  
 RX MEDLINE-97001215; PubMed-8844207;  
 RA Maillet P., Allouais N., Morie L., Delannay J.;  
 RT "Spectrin mutations in hereditary elliptocytosis and hereditary  
 RT spherocytosis.";  
 RL Hum. Mutat. 8:97-107(1996).  
 RN [9]  
 RP VARIANT SER-24.  
 RX MEDLINE-94289716; PubMed-8018926;  
 RA Parquet N., Devaux I., Boulanger L., Galand C., Boylin P.,  
 RA Lecomte M.-C., Dharmy D., Garbacz W.;  
 RT "Identification of three novel spectrin alpha I/74 mutations in  
 RT hereditary elliptocytosis: further support for a triple-stranded  
 RT folding unit model of the spectrin heterodimer contact site.";  
 RL Blood 84:303-308(1994).  
 RN [10]  
 RP VARIANTS CYS-28; HIS-28; LEU-28 AND SER-28.  
 RX MEDLINE-91358728; PubMed-1679439;  
 RA Coetzer T.L., Sahr K., Pirchal J., Blacklock H., Peterson L., Koler R.,  
 RA Doyle J., Manaster J., Palek J.;  
 RT "Four different mutations in codon 28 of alpha spectrin are  
 RT associated with structurally and functionally abnormal spectrin alpha  
 RT I/74 in hereditary elliptocytosis.";  
 RL J. Clin. Invest. 88:743-749(1991).  
 RN [11]  
 RP VARIANTS SER-28 AND ARG-48.  
 RX MEDLINE-91346849; PubMed-1878597;  
 RA Floyd P.B., Gallagher P.G., Valentini L.A., Davis M., Marchesi S.L.,  
 RA Forget B.G.;  
 RT "Heterogeneity of the molecular basis of hereditary  
 RT pyropoikilocytosis and hereditary elliptocytosis associated with  
 RT increased levels of the spectrin alpha I/74-kilodalton tryptic  
 RT peptide.";

RL Blood 78:1364-1372(1991).  
 RN [12]  
 RP VARIANT SER-45.  
 RX MEDLINE-89323468; PubMed-2568862;  
 RA Lecomte M.-C., Garbarz M., Grandchamp B., Feo C., Gautero H.,  
 Deaux I., Bourmer O., Galand C., D'Auriol L., Galibert F.,  
 Sahr K.E., Forget B.G., Bolvin P., Dhery D.;  
 RT "Sp alpha I/78: a mutation of the alpha I spectrin domain in a white  
 RL kindred with HE and HPP phenotypes.";  
 RN Blood 74:1126-1133(1989).  
 RN [13]  
 RP VARIANT PRO-207.  
 RX MEDLINE-92176375; PubMed-1541680;  
 RA Gallagher P.G., Yse W.T., Coetzer T., Lecomte M.-C., Garbarz M.,  
 Zarkowaky H.S., Baruchel A., Ballas S.K., Dhery D., Palek J.,  
 Forget B.G.;  
 RT "A common type of the spectrin alpha I 46-50a-KD peptide abnormality  
 RT in hereditary elliptocytosis and pyropoikilocytosis is associated  
 RT with a mutation distant from the proteolytic cleavage site. Evidence  
 RT for the functional importance of the triple helical model of  
 RL spectrin.";  
 RN J. Clin. Invest. 89:892-898(1992).  
 RN [14]  
 RP VARIANT VAL-1857.  
 RX MEDLINE-93253053; PubMed-8486776;  
 RA Wilmotte R., Marechal J., Morle L., Baklouti F., Philippe N.,  
 Kastally R., Kotula L., Delaunay J., Allosio N.;  
 RT "Low expression allele alpha LEVY of red cell spectrin is associated  
 RT with mutations in exon 40 (alpha V/41 polymorphism) and intron 45 and  
 RT with partial skipping of exon 46.";  
 RN J. Clin. Invest. 91:2091-2096(1993).  
 RN [15]  
 RP VARIANT BARCELONA.  
 RX MEDLINE-93373367; PubMed-8364215;  
 RA dalla Venezia N., Allosio N., Fortsieser A., Denoroy L., Aymerich M.,  
 Vives-Corrons J.L., Besalduch J., Besson I., Delaunay J.;  
 RT "Elliptocytosis associated with the alpha 469 His-->Pro  
 RT mutation in spectrin Barcelona (alpha I/50-46b).";  
 RN Blood 82:1661-1665(1993).  
 RN [16]  
 RP VARIANT CAGLIARI.  
 RX MEDLINE-94043025; PubMed-8226774;  
 RA Sahr K.E., Coetzer T.L., Moy L.S., Derick L.H., Chishti A.H.,  
 Jarolim P., Lorenzo F., Miraglia del Giudice E., Iolascon A.,  
 Gallanelli R.;  
 RT "Spectrin Cagliari: an Ala-->Gly substitution in helix 1 of beta  
 RT spectrin repeat 17 that severely disrupts the structure and self-  
 RT association of the erythrocyte spectrin heterodimer.";  
 RN J. Biol. Chem. 268:22656-22662(1993).  
 RN [17]  
 RP VARIANTS CULO2 AND LYON.  
 RX MEDLINE-90347052; PubMed-2384601;  
 RA Morle L., Roux A.-F., Allosio N., Pothier B., Starck J., Denoroy J.,  
 Morle F., Rudiger R.-C., Forget B.G., Delaunay J., Godet J.;  
 RT "Two elliptocytogenic alpha I/74 variants of the spectrin alpha I  
 RT domain, Spectrin Culo2 (Ggt->Gtt; alpha I 40 Gly-->Val) and  
 RT spectrin Lyon (Cgt-->Ttt; alpha I 43 Leu-->Phe).";  
 RN J. Clin. Invest. 86:548-554(1990).  
 RN [18]  
 RP VARIANT JENDOUBA.  
 RX MEDLINE-92345619; PubMed-1638030;  
 RA Allosio N., Wilmotte R., Morle L., Baklouti F., Marechal J.,  
 Ducluzeau M.-T., Denoroy L., Feo C., Forget B.G., Kastally R.,  
 Delaunay J.;  
 RT "Spectrin Jendouba: an alpha II/31 spectrin variant that is  
 RT associated with elliptocytosis and carries a mutation distant from  
 RT the dimer self-association site.";  
 RN Blood 80:809-815(1992).  
 RN [19]  
 RP VARIANT TUNIS.  
 RX MEDLINE-89323436; PubMed-2568861;  
 RA Morle L., Morle F., Roux A.F., Godet J., Forget B.G., Denoroy L.,  
 Garbarz M., Dhery D., Kastally R., Delaunay J.;

RT "Spectrin Tunis (Sp alpha I/78): an elliptocytogenic variant, is due  
 RT to the CGC->TGC codon change (Arg->Trp) at position 35 of the  
 RT alpha I domain.";  
 RN Blood 74:828-832(1989).  
 RN [20]  
 RP VARIANT GENOVA.  
 RX MEDLINE-94250920; PubMed-8193371;  
 RA Perrotta S., del Giudice E.M., Allosio N., Sclarratta G., Pinto L.,  
 RA Delaunay J., Cutlilo S., Iolascon A.;  
 RT "Mild elliptocytosis associated with the alpha 34 Arg->Trp mutation  
 RT in spectrin Genova (alpha I/74).";  
 RN Blood 83:3346-3349(1994).  
 RN [21]  
 RP VARIANT ANASTASIA.  
 RX MEDLINE-95290423; PubMed-7772539;  
 RA Perrotta S., Iolascon A., de Angelis F., Pagano L., Colonna G.,  
 RA Cutlilo S., del Giudice E.M.;  
 RT "Spectrin Anastasia (alpha I/78): a new spectrin variant (alpha 45  
 RT Arg->Thr) with moderate elliptocytogenic potential.";  
 RN Br. J. Haematol. 89:933-936(1995).  
 CC -1- FUNCTION: SPECTRIN IS THE MAJOR CONSTITUENT OF THE CYTOSKELETAL  
 CC NETWORK UNDERLYING THE ERYTHROCYTE PLASMA MEMBRANE. IT ASSOCIATES  
 CC WITH BAND 4.1 AND ACTIN TO FORM THE CYTOSKELETAL SUPERSTRUCTURE OF  
 CC THE ERYTHROCYTE PLASMA MEMBRANE.  
 CC -1- SUBUNIT: COMPOSED OF NONHOMOLOGOUS CHAINS, ALPHA AND BETA, WHICH  
 CC AGGREGATE SIDE-TO-SIDE IN AN ANTIPARALLEL FASHION TO FORM DIMERS,  
 CC TERRAERS, AND HIGHER POLYMERS.  
 CC -1- DISEASE: HEREDITARY ELLIPTOCYTOSIS (HE) IS A HETEROGENEOUS  
 CC DISORDER CHARACTERIZED BY VARIABLE HEMOLYTIC ANEMIA AND ELLIPTICAL  
 CC RED CELL SHAPE. SEVERAL ABNORMALITIES IN THE MEMBRANE SKELETON  
 CC HAVE BEEN IDENTIFIED IN HE, INCLUDING A NUMBER THAT HAVE BEEN

Query Match 36.4%; Score 47; DB 1; Length 2418;  
 Best Local Similarity 40.9%; Pred. No. 1.7e+02;  
 Matches 9; Conservative 8; Mismatches 5; Indels 0; Gaps 0;

Oy 2 DALEALADGTALQSEBAAVK 23  
 Db 953 DSMKALRNQANACOOQAAAPV 974  
 1::111::1:1:1111:

RESULT 10  
 KPYK\_THEAC  
 ID KPYK\_THEAC STANDARD; PRT; 544 AA.  
 AC P32044;  
 DT 01-JUL-1993 (Rel. 26, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Pyruvate kinase (EC 2.7.1.40) (PK).  
 GN TA0896.  
 OS Thermoplasma acidophilum.  
 OC Archaea; Euryarchaeota; Thermoplasmatales; Thermoplasmaceae;  
 CC Thermoplasma  
 NCBI\_TaxID=2303;  
 OX [1]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN-DSM 1728;  
 RC MEDLINE-20479972; PubMed-11029001;  
 RA Ruepp A., Granel W., Santos-Martinez M.-L., Koretke K.K., Volker C.,  
 RA Mewes H.-W., Frishman D., Stocker S., Lupas A.N., Baumeister W.;  
 RT "The genome sequence of the thermophilic scavenger Thermoplasma  
 RT acidophilum.";  
 RN Nature 407:508-513(2000).  
 RN [2]  
 RP CHARACTERIZATION.  
 RX MEDLINE-93051217; PubMed-1426985;  
 RA Potter S., Fothergill-Gilmore L.A.;  
 RT "Purification and properties of pyruvate kinase from Thermoplasma  
 RT acidophilum.";  
 RN FEBS Microbiol. Lett. 73:235-239(1992).  
 CC -1- CATALYTIC ACTIVITY: ATP + pyruvate = ADP + phosphoenolpyruvate.  
 CC -1- COFACTOR: REQUIRES MAGNESIUM AND POTASSIUM.  
 CC -1- PATHWAY: FINAL STEP IN GLYCOLYSIS.

```

CC -1- SUBUNIT: HOMOTETRAMER.
CC -1- SIMILARITY: BELONGS TO THE PYRUVATE KINASE FAMILY.
CC CAUTION: REF.2 HAS PUBLISHED SOME PARTIAL SEQUENCE. THESE
CC SEQUENCES DO NOT ORIGINATE FROM T. ACIDOPHILUM, RATHER THEY SEEM TO
CC BE CONTAMINATED WITH HUMAN SAMPLES.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: AL45065; CAC12025.1; -.
CC InterPro: IPR001697; Pyruvate_kinase.
CC Pfam: PF00224; PK.1.
CC PRINTS: PR01050; PYRUVTKINASE.
CC PRODOM: PD001009; Pyruvate_kinase.1.
CC PROSITE: PS00110; PYRUVATE_KINASE; FALSE_NEG.
CC Transferase: Kinase; Glycolysis; Magnesium; Complete proteome.
CC ACT_SITE 202 202 BY SIMILARITY.
CC METAL 204 204 MAGNESIUM (BY SIMILARITY).
CC METAL 225 225 MAGNESIUM (BY SIMILARITY).
CC METAL 226 226 MAGNESIUM (BY SIMILARITY).
CC SEQUENCE 544 AA; 59147 MW; 7BD8BCAF4706A5B2 CRC64;

Query Match 36.08; Score 46.5; DB 1; Length 544;
Best Local Similarity 50.08; Pred. No. 42;
Matches 12; Conservative 3; Mismatches 8; Indels 1; Gaps 1;

OY 1 SDALFALADOTDALQ-SEEAAYVK 23
DB 278 SDITNAIDNALMLSESAIGK 301

RESULT 11
PAND_STAAM STANDARD; PRT; 127 AA.
AC 099R40;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Aspartate 1-decarboxylase precursor (EC 4.1.1.11) (Aspartate alpha-
DE decarboxylase).
GN PAND OR SAW2597 OR SA2390.
OS Staphylococcus aureus (Strain M450 / ATCC 700699), and
OS Staphylococcus aureus (Strain N315).
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Staphylococcus.
OX NCBI_TaxID=158878, 158879;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-M450 / ATCC 700699, and N315;
RX MEDLINE=21311952; PubMed=11418146;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Iian J.-O., Ito T.,
RA Kenmori M., Matsuura H., Maruyama A., Murakami H., Hosoyama A.,
RA Mizutani U. Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
RA Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,
RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;
RA Whole genome sequencing of methicillin-resistant Staphylococcus
RT aureus.
RT Lancet 357:1225-1240(2001).
CC -1- CATALYTIC ACTIVITY: L-aspartate -> beta-alanine + CO(2).
CC -1- COFACTOR: Pyruvoyl group. (By similarity).
CC -1- PATHWAY: Pantothenate biosynthesis; second branch.
CC -1- SIMILARITY: BELONGS TO THE PAND FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: AP003365; BAB58759.1; ALT_INIT.
CC EMBL: AP003137; BAB43695.1; ALT_INIT.
CC InterPro: IPR003190; Asp.decarbox.
CC Pfam: PF02261; Asp.decarbox.1.
CC Pantothenate biosynthesis; Lyase; Decarboxylase; Pyruvate; Zymogen;
CC Complete proteome.
CC CHAIN 1 24
FT FT CHAIN 25 127 ASPARTATE 1-DECARBOXYLASE BETA CHAIN (BY
FT MOD_RES 25 25 SIMILARITY).
FT FT MOD_RES 25 25 CONVERTED TO A PYRUVYL GROUP (BY
FT SEQUENCE 127 AA; 14050 MW; F35BF82DB7CA7EA CRC64;

Query Match 35.78; Score 46; DB 1; Length 127;
Best Local Similarity 42.98; Pred. No. 11;
Matches 12; Conservative 5; Mismatches 7; Indels 4; Gaps 1;

OY 1 SDALFALADOTDALQSEEAAYVKADNA 28
DB 30 SDILEAV----DILPNEKVAIVNNNGA 53

RESULT 12
IM44_CAEEL STANDARD; PRT; 425 AA.
AC 002161;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable import inner membrane translocase subunit TIM44,
DE mitochondrial precursor.
GN T09B4.9.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodidae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RL Langston Z., Wohlmann P., Gillam B.;
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: INVOLVED IN PROTEIN IMPORT INTO THE MITOCHONDRION.
CC PROBABLY INVOLVED IN TRANSLLOCATION ACROSS THE INNER MEMBRANE. AS A
CC BINDING PROTEIN REQUIRED FOR DRIVING THE IMPORT OF PREPROTEINS.
CC RECRUITS MITOCHONDRIAL HSP70 TO DRIVE PROTEIN TRANSLLOCATION INTO
CC THE MATRIX USING ATP AS AN ENERGY SOURCE (BY SIMILARITY).
CC -1- SUBUNIT: FORMS PART OF THE RECEPTOR COMPLEX THAT CONSISTS OF AT
CC LEAST 3 DIFFERENT PROTEINS (TIM17, TIM23, TIM44) (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Mitochondrial inner membrane (Potential).
CC -1- SIMILARITY: BELONGS TO THE TIM44 FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: U97405; AAB53011.1; -.
CC WormPep: T09B4.9; CE13473.
CC Mitochondrion; Inner membrane; Transport; Protein transport;
CC TRANSLOCATION; Transl peptide.
CC TRANSIT 1 425
FT FT CHAIN ? 425 MITOCHONDRION.
FT SEQUENCE 425 AA; 49398 MW; 203DFBD614E09F8 CRC64;

```

Query Match 35.7%; Score 46; DB 1; Length 425;  
 Best Local Similarity 41.7%; Pred. No. 38;  
 Matches 10; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

QY 5 EALADOTDALOSEAAVYKADNNA 28  
 DB 182 EALTKRTDGFDEKERVVANDSA 205

RESULT 13  
 HTR2\_HALNI STANDARD: PRT: 763 AA.

AC Q9HP81; 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Sensory rhodopsin II transducer (HTR-II) (Methyl-accepting phototaxis protein II) (MPP-II).  
 GN HTR2 OR VNG1765G.  
 OS Halobacterium sp. (strain NRC-1).  
 OC Archaea: Euryarchaeota; Halobacteriales; Halobacteriaceae;  
 CC Halobacterium  
 CC NCBI\_TaxID=64091;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20504483; PubMed=11016950;  
 RA Ng W.V., Kennedy S.P., Mahalax G.G., Bergquist B., Pan M., Shukla H.D., Lasky S.R., Baliga S.S., Thorsson V., Schroga J., Swartzell S., Welt D., Hall J., Dahl T.A., Welt R., Goo T.A., Leihauer B., Keller K., Cruz R., Danson M.J., Hough D.W., Maddocks D.G., Jablonski P.E., Krebs M.P., Angelino C.M., Dale H., Aisenbarger T.A., Peck R.F., Pohlschoder M., Spudis J.L., Jung K.-H., Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D., Eberhart H., Lowe T.M., Liang P., Riley M., Hood L., Datta S.S.;  
 RA "Genome sequence of Halobacterium species NRC-1";  
 RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).  
 CC -1- FUNCTION: TRANSDUCES SIGNALS FROM THE PHOTOTAXIS RECEPTOR SENSORY RHODOPSIN II (SR-II) TO THE FLAGELLAR MOTOR. RESPONDS TO LIGHT CHANGES THROUGH THE VARIATION OF THE LEVEL OF METHYLATION (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE CHEMOTAXIS SENSORY TRANSDUCERS FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sdb.ch/announce/> or send an email to [license@sdb-sib.ch](mailto:license@sdb-sib.ch)).  
 CC EMBL: AE005080; AAC19989.1; -  
 DR InterPro: IPR004089; Chemotaxis\_transducer.  
 DR InterPro: IPR003660; HAMF.  
 DR Pfam: PF00672; HAMF; 1.  
 DR Pfam: PF00015; MCPsignal; 1.  
 DR SMART: SM00304; HAMF; 2.  
 DR SMART: SM00283; MA; 1.  
 KW Transducer; Photoreceptor; Transmembrane; Methylation;  
 KW Complete proteome.  
 FT INIT\_MET 0 BY SIMILARITY.  
 FT DOMAIN 1 15 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 16 36 POTENTIAL.  
 FT DOMAIN 37 277 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 278 297 POTENTIAL.  
 FT DOMAIN 298 763 CYTOPLASMIC (POTENTIAL).  
 SQ SEQUENCE 763 AA; 78911 MW; CF7ABFF04DFE309A CRC64;

Query Match 35.7%; Score 46; DB 1; Length 763;  
 Best Local Similarity 52.4%; Pred. No. 70;  
 Matches 11; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY 1 SPALALADOTDALOSEAAV 21

DB 550 ADDAEAAADAMDALDSEMAI 570

RESULT 14  
 HTR2\_HALSA STANDARD: PRT: 764 AA.

AC P71410; 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Sensory rhodopsin II transducer (HTR-II) (Methyl-accepting phototaxis protein II) (MPP-II).  
 GN HTR2 OR HTRII.  
 OS Halobacterium salinarum.  
 OC Archaea: Euryarchaeota; Halobacteriales; Halobacteriaceae;  
 CC Halobacterium  
 CC NCBI\_TaxID=2242;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FLX15;  
 RX MEDLINE=96323203; PubMed=8710852;  
 RA Zhang W., Brooun A., Mueller M.M., Alam M.;  
 RT "The primary structures of the Archaeon Halobacterium salinarum blue light receptor sensory rhodopsin II and its transducer, a methyl-accepting protein.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 93:8230-8235(1996).  
 CC -1- FUNCTION: TRANSDUCES SIGNALS FROM THE PHOTOTAXIS RECEPTOR SENSORY RHODOPSIN II (SR-II) TO THE FLAGELLAR MOTOR. RESPONDS TO LIGHT CHANGES THROUGH THE VARIATION OF THE LEVEL OF METHYLATION.  
 CC -1- SIMILARITY: BELONGS TO THE CHEMOTAXIS SENSORY TRANSDUCERS FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sdb.ch/announce/> or send an email to [license@sdb-sib.ch](mailto:license@sdb-sib.ch)).  
 CC EMBL: U62676; AAC4369.1; -  
 DR HSP: P02942; 1007.  
 DR InterPro: IPR004089; Chemotaxis\_transducer.  
 DR InterPro: IPR003660; HAMF.  
 DR Pfam: PF00672; HAMF; 1.  
 DR Pfam: PF00015; MCPsignal; 1.  
 DR SMART: SM00304; HAMF; 2.  
 DR SMART: SM00283; MA; 1.  
 KW Transducer; Photoreceptor; Transmembrane; Methylation;  
 KW INIT\_MET 0 BY SIMILARITY.  
 FT DOMAIN 1 15 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 16 36 POTENTIAL.  
 FT DOMAIN 37 277 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 278 298 POTENTIAL.  
 FT DOMAIN 299 764 CYTOPLASMIC (POTENTIAL).  
 SQ SEQUENCE 764 AA; 79187 MW; 1E0D7B4E460FC588 CRC64;

Query Match 35.7%; Score 46; DB 1; Length 764;  
 Best Local Similarity 52.4%; Pred. No. 70;  
 Matches 11; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY 1 SPALALADOTDALOSEAAV 21

DB 551 ADDAEAAADAMDALDSEMAI 571

RESULT 15  
 YC88\_MYCPN STANDARD: PRT: 787 AA.

AC P75489; 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)



```

DE Hypothetical lipoprotein MPN288 precursor (A65_orf7870).
GN MPN288 OR MP547.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.
RN NCBI_TaxID=2104;
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 29342 / M129.
RA MEDLINE-97103885; PubMed-8948633;
RA Himmelfeich R., Hilbert H., Plagens H., Plirk E., Li B.-C.,
RA Hermann R.;
RT *Complete sequence analysis of the genome of the bacterium Mycoplasma
RT pneumoniae.
RL Nucleic Acids Res. 24:4420-4449(1996).
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
CC (Potential).
CC -1- SIMILARITY: BELONGS TO THE MG185 / MG260 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE000053; AAB96195.1;
KW Hypothetical protein; Lipoprotein; Membrane; Signal;
KW Complete proteome.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 787 HYPOTHETICAL LIPOPROTEIN MPN288.
FT LIPID 23 23 N-ACYL DIGLYCERIDE (POTENTIAL).
SQ SEQUENCE 787 AA; 86889 MW; FFA6237DD18D3A05 CRC64;

Query Match
Best Local Similarity 35.7%; Score 46; DB 1; Length 787;
Matches 10; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

OY 3 ALEALADQTDALQSEEAAYKAA 24
DB 599 ATEVELEQSTQSDRAIVLPA 620

RESULT 16
Y182_STRCO STANDARD: PRT; 1000 AA.
AC 09FC14; Q9K4B4;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein 2SC3B6.28/SC7E4.01.
GN 2SC3B6.28 OR SC7E4.01.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Seeger K.J., Harris D., Cerdano A.M., Parkhill J., Barrel B.G.,
RA Rajandream M.A.;
RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: BELONGS TO THE UPF0182 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

CC -----
DR EMBL: AL390968; CAC01334.1;
DR EMBL: AL359214; CAB94589.1;
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 26 48 POTENTIAL.
FT TRANSMEM 70 92 POTENTIAL.
FT TRANSMEM 121 143 POTENTIAL.
FT TRANSMEM 177 199 POTENTIAL.
FT TRANSMEM 220 237 POTENTIAL.
FT TRANSMEM 267 289 POTENTIAL.
FT TRANSMEM 296 318 POTENTIAL.
SQ SEQUENCE 1000 AA; 109554 MW; C30AB349CA9E410 CRC64;

Query Match
Best Local Similarity 35.7%; Score 46; DB 1; Length 1000;
Matches 12; Conservative 1; Mismatches 11; Indels 0; Gaps 0;

OY 5 EALADQTDALQSEEAAYKADNAA 28
DB 936 EAOKDLEALQRAEDQAQAKDQCA 959

RESULT 17
MAX_BRARE STANDARD: PRT; 165 AA.
AC P52161;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE MAX protein.
GN MAX.
OS Brachydanio rerio (Zebrafish) (Zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC Cypriniformes; Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE-93233639; PubMed-8474440;
RA Schreiber-Agus N., Horner J., Torres R., Chiu F.-C., Depinho R.A.;
RT "Zebra fish myc family and max genes: differential expression and
RT oncogenic activity throughout vertebrate evolution.";
RL Mol. Cell. Biol. 13:2765-2775(1993).
CC -1- FUNCTION: TRANSCRIPTION REGULATOR. FORMS A SPOUNGE-SPECIFIC
CC DNA-BINDING PROTEIN COMPLEX WITH MYC OR MAD WHICH RECOGNIZES
CC THE CORE SEQUENCE 5'-CAC(GA)TG-3'. THE MYC-MAX COMPLEX IS A
CC TRANSCRIPTIONAL ACTIVATOR, WHEREAS THE MAD-MAX COMPLEX IS A
CC REPRESSOR (BY SIMILARITY).
CC -1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER
CC BHLH PROTEIN. BINDS DNA AS A HETERODIMER WITH MYC OR MAD.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A LONG FORM (SHOWN HERE) AND A
CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: HIGH LEVELS ARE SEEN IN THE KIDNEY, GILLS AND
CC UTERUS. IT IS ALSO FOUND IN THE BRAIN AND HEART.
CC -1- PTM: PHOSPHORYLATED (PROBABLE).
CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC TRANSCRIPTION FACTORS. BHLH-ZIP SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L11711; AAA02483.1;
DR HSSP: P25912; 1HL0.
DR ZFIN: ZDB-GENE-990415-152; max.
DR InterPro: IPR003015; HLH_Myc.
DR InterPro: IPR001092; HLH_dim.
DR Pfam: PF00010; HLH_1.

```

```

DR SMART: SM00353; HLM; 1.
DR PROSITE: PS00038; HELIX-LOOP-HELIX; 1.
KW Nuclear protein; DNA-binding; Transcription regulation;
KW Alternative splicing; Phosphorylation.
FT DNA_BIND 24 36 BASIC DOMAIN.
FT DOMAIN 39 80 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).
FT DOMAIN 86 107 LEUCINE-ZIPPER (BY SIMILARITY).
FT VARSPLIC 13 21 MISSING (IN SHORT ISOFORM).
SQ SEQUENCE 165 AA; 18728 MW; 227A31C7C415821A CRC64;

Query Match 34.9%; Score 45; DB 1; Length 165;
Best Local Similarity 34.6%; Pred. NO. 19;
Matches 9; Conservative 6; Mismatches 11; Indels 0; Gaps 0;

OY 2 DALEALADQTDALQSEEAAYVKADNA 27
DB 41 DSFHSRLDSVPALQCKQSIKQASRA 66

RESULT 18
KARL_YEAST STANDARD: PRT: 433 AA.
AC P11927;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Cell division control protein KARL.
CN KARL OR YNL188W OR N1611.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycos.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87159524; PubMed=3030557;
RA Rose M.D., Fink G.R.;
RT "KARL, a gene required for function of both Intracellular and
RT extracellular microtubules in yeast.";
RL Cell 48:1047-1060(1987).
RN [2]
RP SEQUENCE FROM N.A.
RA Obermaler B., Piravandi E., Rinke M., Domdey H.;
RL Submitted (May-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP FUNCTION.
RX MEDLINE=94350199; PubMed=8070654;
RA Vallen E.A., Ho W., Winey M., Rose M.D.;
RT "Genetic interactions between CDC31 and KARL, two genes required for
RT duplication of the microtubule organizing center in Saccharomyces
RT cerevisiae.";
RL Genetics 137:407-422(1994).
RN [4]
RP FUNCTION.
RC STRAIN=5288C;
RX MEDLINE=99181541; PubMed=7876310;
RA Spang A., Courtney I., Grein K., Matzner M., Schebel E.;
RT "The Cdc31p-binding protein Kar1p is a component of the half bridge
RT of the yeast spindle pole body.";
RL J. Cell Biol. 128:863-877(1995).
CC -1- FUNCTION: KARL IS REQUIRED FOR FUNCTION OF BOTH INTRANUCLEAR AND
CC EXTRANUCLEAR MICROTUBULES. KARL HELPS LOCALIZE CDC31 TO THE
CC SPINDLE POLE BODY (SPB), CDC31 THEN INITIATES SPB DUPLICATION VIA
CC INTERACTION WITH A DOWNSTREAM EFFECTOR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M15683; AAA34716.1; -.

```

```

DR EMBL: Z71464; CAA96082.1; -.
DR PIR: A26455; A26455.
DR SGD: S0005132; KARL.
KW Cell division; Microtubules; Mitosis.
FT CONFLICT 199 199 A -> V (IN REF. 1).
SQ SEQUENCE 433 AA; 50653 MW; D65331B3B2B068E4 CRC64;

Query Match 34.9%; Score 45; DB 1; Length 433;
Best Local Similarity 48.0%; Pred. NO. 53;
Matches 12; Conservative 3; Mismatches 8; Indels 2; Gaps 1;

OY 4 LEALADQTDALQSEEAAYVKADNA 28
DB 360 IELLKDDTDS--KEKKRYVTNDNAA 382

RESULT 19
KCC4_RAT STANDARD: PRT: 474 AA.
AC P13234;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Calcium/calmodulin-dependent protein kinase type IV catalytic chain
DE (EC 2.7.1.123) (CAM kinase-GR) (CAMK IV) [contains: Caldesmonin].
CN CAMK4.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=91288548; PubMed=1648230;
RA Ohmstede C.-A., Bland M.M., Merrill B.M., Sahyoun N.;
RT "Relationship of genes encoding Ca2+/calmodulin-dependent protein
RT kinase Gr and caldesmonin: a gene within a gene.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:5784-5788(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=91304387; PubMed=1649385;
RA Means A.R., Cruzalegui F., Lemagueresse B., Needleman D.S.,
RA Slaughter G.R., Ono T.;
RT "A novel Ca2+/calmodulin-dependent protein kinase and a male germ
RT cell-specific calmodulin-binding protein are derived from the same
RT gene.";
RL Mol. Cell. Biol. 11:3960-3971(1991).
RN [3]
RP SEQUENCE OF 250-474 FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Brain;
RX MEDLINE=89174647; PubMed=2538431;
RA Ohmstede C.-A., Jensen K.F., Sahyoun N.;
RT "Ca2+/calmodulin-dependent protein kinase enriched in cerebellar
RT granule cells. Identification of a novel neuronal
RT calmodulin-dependent protein kinase.";
RL J. Biol. Chem. 264:5866-5875(1989).
RN [4]
RP SEQUENCE OF 306-474 FROM N.A., AND SEQUENCE OF 335-361.
RC STRAIN=SPRAGUE-DAWLEY;
RX MEDLINE=89123272; PubMed=2914893;
RA Ono T., Slaughter G.R., Cook R.G., Means A.R.;
RT "Molecular cloning sequence and distribution of rat caldesmonin, a
RT high affinity calmodulin-binding protein.";
RL J. Biol. Chem. 264:2081-2087(1989).
CC -1- FUNCTION: CAM KINASE GR IS A NEURONAL-SPECIFIC PROTEIN KINASE,
CC ENRICHED IN CEREBELLAR GRANULE CELLS.
CC -----
CC -1- FUNCTION: CALSPERMIN IS A HEAT-STABLE, ACIDIC, CALMODULIN-BINDING
CC PROTEIN.
CC -1- CATALYTIC ACTIVITY: ATP + protein = ADP + O-phosphoprotein.
CC -1- ALTERNATIVE PRODUCTS: CALCIUM/CALMODULIN-DEPENDENT PROTEIN KINASE
CC TYPE IV CATALYTIC CHAIN AND CALSPERMIN ARE OBTAINED BY ALTERNATIVE
CC SPLICING.
CC -----
CC -1- TISSUE SPECIFICITY: CALSPERMIN IS PREDOMINANTLY FOUND IN MAMMALIAN
CC TESTIS AND BRAIN.

```

CC -1- PTM: THE N-TERMINAL OF CALSPERMIN IS BLOCKED.  
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 CC CAMK SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: M63334; AAA40865.1; -  
 CC EMBL: M74488; AAA40845.1; ALT\_SEQ.  
 CC EMBL: M64757; AAA40856.1; -  
 CC EMBL: M64757; AAA40857.1; -  
 CC EMBL: J04460; AAA41867.1; -  
 CC EMBL: J04446; AAA40990.1; -  
 CC PIR: A41103; TVRTCA.  
 CC HSP: P00518; 1PHK.  
 DR InterPro: IPR000719; Euk\_pkinase.  
 DR InterPro: IPR002290; Ser\_thr\_kinase.  
 DR Pfam: PF00069; pkinase; 1.  
 DR SMART: SM00220; S\_TKC; 1.  
 DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE: PS00108; PROTEIN\_KINASE\_ST; 1.  
 DR PROSITE: PS50011; PROTEIN\_KINASE\_DOM; 1.  
 DR Transferrase: Serine/threonine-protein kinase: ATP-binding;  
 KW Calmodulin-binding; Testis; Alternative splicing.  
 FT CHAIN 306 474  
 FT DOMAIN 42 296  
 FT NP\_BIND 48 56  
 FT BINDING 71 71  
 FT ACT\_SITE 160 160  
 FT DOMAIN 338 337  
 FT DOMAIN 393 399  
 FT DOMAIN 403 413  
 FT CONFLICT 372 372 I -> M (IN REF. 2 AND 4).  
 SQ SEQUENCE 474 AA: 53133 MW: 56F71AC5644DE23 CRC64;  
 Query Match 34.9%; Score 45; DB 1; Length 474;  
 Best Local Similarity 58.8%; Pred. No. 58;  
 Matches 10; Conservative 3; Mismatches 4; Indels 0; Gaps 0;  
 QY 2 DALEALADOTDAQSEF 18  
 Db 378 EAAKAAADETMKLQSEE 394  
 RESULT 20  
 PPBH\_PSEAE STANDARD; PRT; 476 AA.  
 AC P35483; Q9HYU7;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Alkaline phosphatase H precursor (EC 3.1.3.1) (High molecular weight  
 DE phosphatase) (H-AP).  
 GN PHO A OR PA3296.  
 OS Pseudomonas aeruginosa.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 OC Pseudomonas.  
 CC NCBI\_TaxID=287;  
 RN NCBI\_TaxID=287;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 15692 / PA01;  
 RX MEDLINE=20437337; PubMed=10984043;  
 RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,  
 RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,  
 RA Gardner R.L., Goltier S.N., Tolentino E., Westbrook-Wadman S., Yuan Y.,  
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,  
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
 RA Reiter J., Sater M.H., Hancock R.E.W., Lory S., Olson M.V.;

RT \*Complete genome sequence of Pseudomonas aeruginosa PA01, an  
 RT opportunistic pathogen\*;  
 RT Nature 406:959-964(2000).  
 RN (2)  
 RP SEQUENCE OF 27-45.  
 RC STRAIN=H103;  
 RX MEDLINE=93202452; PubMed=8454193;  
 RA Tan A.S.P., Morobec E.A.;  
 RT Isolation and characterization of two immunochemically distinct  
 RT alkaline phosphatases from Pseudomonas aeruginosa\*;  
 RT FEMS Microbiol. Lett. 106:281-286(1993).  
 CC -1- FUNCTION: HAS ONLY A PHOSPHOMONESTERASE ACTIVITY.  
 CC -1- CATALYTIC ACTIVITY: An orthophosphoric monoester + H(2)O -> an  
 CC alcohol + phosphate.  
 CC -1- COPFACTOR: METALLOENZYME CONTAINING TWO ZINC ATOMS AND A MAGNESIUM  
 CC ION.  
 CC -1- SUBCELLULAR LOCATION: SECRETED AND PERIPLASMIC.  
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED CONSTITUTIVELY.  
 CC -1- SIMILARITY: BELONGS TO THE ALKALINE PHOSPHATASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AE004751; AAG06684.1; -  
 DR InterPro: IPR001952; Alk\_phosphatase.  
 DR Pfam: PF00245; alk\_phosphatase; 1.  
 DR PRINTS: PR00113; ALKPHPHASE.  
 DR PRODOM: PD001668; Alk-phosphatase; 1.  
 DR SMART: SM00098; alkPec; 1.  
 DR PROSITE: PS00123; ALKALINE\_PHOSPHATASE; 1.  
 DR Hydrolyase; Zinc; Magnesium; Periplasmic; Signal; Complete proteome.  
 FT SIGNAL 1 26  
 FT CHAIN 27 476  
 FT CONFLICT 27 27 R -> Q (IN REF. 2).  
 FT CONFLICT 39 39 R -> A (IN REF. 2).  
 FT CONFLICT 44 45 EY -> VR (IN REF. 2).  
 SQ SEQUENCE 476 AA: 50408 MW: A3059B30BBEA809C CRC64;  
 Query Match 34.9%; Score 45; DB 1; Length 476;  
 Best Local Similarity 43.3%; Pred. No. 59;  
 Matches 13; Conservative 5; Mismatches 10; Indels 2; Gaps 1;  
 QY 1 SDALEALADOTDA--LQSEBAVVKADNAA 28  
 Db 330 SKAIELLKDNPNNGFFLQVEGASIDKQDHAA 359  
 RESULT 21  
 RSP3\_CHLRE STANDARD; PRT; 516 AA.  
 ID RSP3\_CHLRE  
 AC P12759;  
 DT 01-OCT-1989 (Rel. 12, Created)  
 DT 01-OCT-1989 (Rel. 12, Last sequence update)  
 DT 01-OCT-1993 (Rel. 27, Last annotation update)  
 DE Radial spoke protein 3.  
 GN RSP3  
 OS Chlamydomonas reinhardtii.  
 OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;  
 OC Chlamydomonadaceae; Chlamydomonas.  
 CC NCBI\_TaxID=3055;  
 RN NCBI\_TaxID=3055;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=21GR.  
 RX MEDLINE=89308863; PubMed=2745550;  
 RA Williams B.D., Velleca M.A., Curry A.M., Rosenbaum J.L.;  
 RT Molecular cloning and sequence analysis of the Chlamydomonas gene  
 RT coding for radial spoke protein 3: flagellar mutation pf-14 is an  
 RT ochre allele.\*;

```

RL J. Cell Biol. 109:235-245(1989).
CC -1- FUNCTION: PROTEIN 3 MAY ATTACH THE RADIAL SPOKE TO THE OUTER
CC DOUBLET MICROTUBULE OR IS REQUIRED TO FORM A STABLE SPOKE
CC STRUCTURE.
CC -1- FUNCTION: FLAGELLAR RADIAL SPOKES CONTRIBUTE TO THE REGULATION
CC OF DINEIN ARM ACTIVITY AND THUS THE PATTERN OF FLAGELLAR BENDING.
CC THEY CONSIST OF A THIN STALK, WHICH IS ATTACHED TO THE A SUBFIBER
CC OF THE OUTER DOUBLET MICROTUBULE, AND A BULBOUS HEAD, WHICH IS
CC ATTACHED TO THE STALK AND APPEARS TO INTERACT WITH THE
CC PROJECTIONS FROM THE CENTRAL PAIR OF MICROTUBULES.
CC -1- SUBCELLULAR LOCATION: RADIAL SPOKE.
CC -1- PTM: PROTEIN 3 IS ONE OF THE 5 RADIAL SPOKE PROTEINS THAT ARE
CC PHOSPHORYLATED.
CC -1- PTM: THE ISOFORM PROTEIN 3A MIGHT BE SIMPLY THE UNPHOSPHORYLATED
CC POLYPEPTIDE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; X14549; CAA32685.1; -.
DR PIR; S05962; S05962.
DR PIR; A31270; A31270.
KW Flagella; Phosphorylation.
SQ SEQUENCE 516 AA; 56784 MW; E928B8F002C48093F CRC64;

Query Match 34.5%; Score 44.5; DB 1; Length 516;
Best Local Similarity 51.7%; Pred. NO. 75;
Matches 15; Conservative 4; Mismatches 7; Indels 3; Gaps 2;

Oy 1 SDALALADOTDALOSEFAVVK--ADNA 27
Db 369 ADAAVAAAELTA-QAEBAANKMEADKA 396

RESULT 22
HTR6_HALN1
ID HTR6_HALN1 STANDARD: PRT; 778 AA.
AC Q9HR92;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Halobacterial transducer protein VI.
GN HTR4 OR HRPVI OR VNC0806G.
OS Halobacterium sp. (strain NRC-1).
OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;
CC Halobacterium.
OX NCBI_TaxID=64091;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20504483; PubMed=11016950;
RA Ng W.V., Kennedy S.P., Mahairas G.G., Herquist B., Pan M.,
RA Shukla H.D., Lasky S.R., Balliga N.S., Thorsson V., Sprogon J.,
RA Swartzell S., Weir D., Hall J., Dahl T.A., Weir R., Goo Y.A.,
RA Lettner B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA Maddocks D.G., Jabloner P.E., Krebs M.P., Angevine C.M., Dale H.,
RA Isenbarger T.A., Peck R.F., Pohlschroder M., Spudis J.L., Jung K.-H.,
RA Altm M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA Eubardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.;
RT "Genome sequence of Halobacterium species NRC-1."
RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
CC -1- FUNCTION: POTENTIALLY INVOLVED IN CHEMO- OR PHOTOTACTIC SIGNAL
CC TRANSDUCTION (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CHEMOTAXIS SENSORY TRANSDUCERS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AE005022; AAG19266.1; -.
DR InterPro; IPR004089; Chemotaxis_transducer.
DR InterPro; IPR003660; HAMF.
DR InterPro; IPR004090; Mc_chemotaxis.
DR Pfam; PF00672; HAMF; 1.
DR Pfam; PF00015; MCPsignal; 1.
DR PRINTS; PR00260; CHEMTRNSDUCR.
DR SMART; SM00304; HAMF; 2.
DR SMART; SM00283; MA; 1.
KW Transducer; Transmembrane; Complete proteome.
FT DOMAIN 1 26 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 27 47 POTENTIAL.
FT DOMAIN 48 296 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 297 317 POTENTIAL.
FT DOMAIN 318 778 CYTOPLASMIC (POTENTIAL).
SQ SEQUENCE 778 AA; 82076 MW; D752278727AAFA79 CRC64;

Query Match 34.5%; Score 44.5; DB 1; Length 778;
Best Local Similarity 39.3%; Pred. NO. 11e+02;
Matches 11; Conservative 4; Mismatches 12; Indels 1; Gaps 1;

Oy 2 DALEALADOTDALOSE-EAAVVKADNAA 28
Db 511 DDLAAVASETDEMSATIEVVAASADVA 538

RESULT 23
HTR6_HALSA
ID HTR6_HALSA STANDARD: PRT; 778 AA.
AC Q48319;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Halobacterial transducer protein VI.
GN HTR4 OR HRPVI.
OS Halobacterium salinarum.
OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;
CC Halobacterium.
OX NCBI_TaxID=2242;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=S9;
RX MEDLINE=96275896; PubMed=8674984;
RA Rudolph J., Nordmann B., Storch K.F., Gruenberg H., Rodewald K.,
RA Oesterhelt D.;
RT "A family of halobacterial transducer proteins."
RL FEMS Microbiol. Lett. 139:161-168(1996).
CC -1- FUNCTION: POTENTIALLY INVOLVED IN CHEMO- OR PHOTOTACTIC SIGNAL
CC TRANSDUCTION.
CC -1- SIMILARITY: BELONGS TO THE CHEMOTAXIS SENSORY TRANSDUCERS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; X95590; CAA64843.1; -.
DR InterPro; IPR004089; Chemotaxis_transducer.
DR InterPro; IPR003660; HAMF.
DR Pfam; PF00672; HAMF; 1.
DR Pfam; PF00015; MCPsignal; 1.
DR SMART; SM00304; HAMF; 2.
DR SMART; SM00283; MA; 1.
KW Transducer; Transmembrane.
FT DOMAIN 1 26 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 27 47 POTENTIAL.

```

FT DOMAIN 48 296 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 297 317 POTENTIAL.  
 FT DOMAIN 318 778 CYTOPLASMIC (POTENTIAL).  
 SO SEQUENCE 778 AA: 82077 MW: 134C7DF0A3334CD CRC64;

Query Match 34.5%; Score 44.5; DB 1; Length 778;  
 Best Local Similarity 39.3%; Pred. No. 1.1e+02;  
 Matches 11; Conservative 4; Mismatches 12; Indels 1; Gaps 1;

QY 2 DALEALADOTDALQSE-EAAVAKADNNA 28  
 DB 511 DOLEAVASETDEMSATIEEVAASADQVA 538

RESULT 24  
 ID ABCX\_ANTSP STANDARD; PRT; 251 AA.  
 AC 002856;  
 DT 01-JUL-1993 (Rel. 26, Created)  
 DT 01-JUL-1993 (Rel. 26, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE Probable ATP-dependent transporter ycf16.  
 GN YCF16.  
 OS Anthlamnion sp.  
 OG Chloroplast.  
 OC Eukaryota; Rhodophyta; Florideophyceae; Ceramiales; Ceramiales; Ceramiales;  
 OC Anthlamnion.  
 OX NCBI\_TaxID=2767;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=LB 95.79;  
 RX MEDLINE=93021132; PubMed=1404401;  
 RA Kostzawa M., Zetsche K.;  
 RT "Large ATP synthase operon of the red alga Anthlamnion sp. resembles the corresponding operon in cyanobacteria.";  
 RL J. Mol. Biol. 227:961-970(1992).  
 CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. YCF16 SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: X63382; CAA44985.1; -;  
 DR PIR: S37635; S37635.  
 DR Mendel: 6183; ANTSP.ycf16.1.  
 DR InterPro: IPR003439; ABC\_transporter.  
 DR Pfam: PF00005; ABC\_tran.1.  
 DR PROSITE: PS00211; ABC\_TRANSPORTER.1.  
 KM ATP-binding; Transport: Chloroplast.  
 FT NP\_BIND 39 46  
 SO SEQUENCE 251 AA: 28205 MW: 3707DD1736A73419 CRC64;

Query Match 34.1%; Score 44; DB 1; Length 251;  
 Best Local Similarity 33.3%; Pred. No. 41;  
 Matches 7; Conservative 10; Mismatches 4; Indels 0; Gaps 0;

QY 2 DALEALADOTDALQSE-EAAV 22  
 DB 184 DALTIKAIKINSIKTOENSI 204

RESULT 25  
 ID CHEA\_LISIN STANDARD; PRT; 618 AA.  
 AC 092DW2;  
 DT 01-MAR-2002 (Rel. 41, Created)  
 DT 01-MAR-2002 (Rel. 41, Last sequence update)

DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Chemotaxis protein cheA (EC 2.7.3.-).  
 GN CHEA OR LIN0700.  
 OS Listeria innocua.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
 CC Bacillus/staphylococcus group; Listeria.  
 OX NCBI\_TaxID=1642;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RC STRAIN=CLIP 11262 / Serovar 6a;  
 RX MEDLINE=21537279; PubMed=11679669;  
 RA Glaser P., Frangeul L., Buchrieser C., Rusnak C., Amend A.,  
 RA Baquero F., Berche P., Bloecher H., Brandt P., Chakraborty T.,  
 RA Charbit A., Chetouani F., Couve E., de Daruvar A., Deloux P.,  
 RA Doman E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurel O.,  
 RA Entian K.-D., Esli H., Garcia-del Portillo F., Garrido P.,  
 RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,  
 RA Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurapat G.,  
 RA Medueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,  
 RA Nordstiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,  
 RA Rammel B., Rose M., Schlueter T., Simoes N., Tierrez A.,  
 RA Vazquez-Boland J.-A., Voss H., Weiland J., Cossart P.;  
 RT "Comparative genomics of Listeria species.";  
 RL Science 294:849-852(2001).  
 CC -1- FUNCTION: INVOLVED IN THE TRANSMISSION OF SENSORY SIGNALS FROM THE CHEMORECEPTORS TO THE FLAGELLAR MOTORS. CHEA IS AUTOPHOSPHORYLATED. IT CAN TRANSFER ITS PHOSPHATE GROUP TO EITHER CHEB OR CHEY (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).  
 CC -1- SIMILARITY: TO OTHER PROKARYOTIC SENSORY TRANSDUCTION HISTIDINE KINASES.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL: AL596166; CAC95932.1; -;  
 DR Listlist: LIN00700; -;  
 KM Sensory transduction; Transference; Kinase; Phosphorylation;  
 KW Chemotaxis; Complete proteome  
 FT MOD\_RSS 45 45 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 SO SEQUENCE 618 AA: 68668 MW: 4302DAA5BDAB458 CRC64;

Query Match 34.1%; Score 44; DB 1; Length 618;  
 Best Local Similarity 41.7%; Pred. No. 1.1e+02;  
 Matches 15; Conservative 4; Mismatches 5; Indels 12; Gaps 2;

QY 4 LEALHPQETDFAVEKTYRAIIEEALIKAVRA 27  
 DB 119 LEALHPQETDFAVEKTYRAIIEEALIKAVRA 154

RESULT 26  
 ID CHEA\_LISMO STANDARD; PRT; 618 AA.  
 AC 048768;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-MAR-2002 (Rel. 41, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Chemotaxis protein cheA (EC 2.7.3.-).  
 GN CHEA OR LM00692.  
 OS Listeria monocytogenes.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
 CC Bacillus/staphylococcus group; Listeria.  
 OX NCBI\_TaxID=1639;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=12067;  
 RX MEDLINE=95102110; PubMed=7803815;



```

DR EMBL: Z71261; CAA95848.1; JOINED.
DR EMBL: Z71261; CAA95806.1; -.
DR EMBL: Z71266; CAA95806.1; JOINED.
DR PIR: S02772; MKRWL.
DR HSSP: P08799; 1MND.
DR WormPep: R06C7.10; CE06253.
DR InterPro: IPR004009; Myosin_N.
DR InterPro: IPR002928; Myosin_tail.
DR InterPro: IPR001609; Myosin_head.
DR Pfam: PF00063; myosin_head.1.
DR Pfam: PF02736; Myosin_N.1.
DR Pfam: PF01576; Myosin_tail.1.
DR PRINTS: PR00193; MYOSINHEAVY.
DR ProDom: PD000355; myosin_head.1.
DR SMART: SM00242; Myosin.1.
KW Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
KW ATP-binding; Methylation; Alkylation; Multigene family.
FT DOMAIN 1 845 MYOSIN HEAD-LIKE.
FT DOMAIN 846 1938 ROD-LIKE TAIL (S2 AND LAM DOMAINS).
FT DOMAIN 1171 1938 ALPHA-HELICAL TAILPIECE (SHORT S2).
FT NP_BIND 177 184 COILED COIL (POTENTIAL).
FT DOMAIN 660 682 ATP (BY SIMILARITY).
FT DOMAIN 764 778 ACTIN-BINDING.
FT MOD_RES 128 128 METHYLATION (TR1-).
FT MOD_RES 700 700 ALKYLATION (SH-1).
FT MOD_RES 710 710 ALKYLATION (SH-2).
FT CONFLICT 94 94 F -> E (IN REF. 2).
FT CONFLICT 98 98 F -> R (IN REF. 2).
FT CONFLICT 377 377 V -> D (IN REF. 4).
FT CONFLICT 389 390 V -> D (IN REF. 4).
FT CONFLICT 391 391 V -> D (IN REF. 4).
FT CONFLICT 408 408 W -> N (IN REF. 2).
FT CONFLICT 474 474 W -> G (IN REF. 2).
FT CONFLICT 577 577 L -> F (IN REF. 4).
FT CONFLICT 681 681 I -> N (IN REF. 4).
FT CONFLICT 1373 1373 S -> D (IN REF. 2).
FT CONFLICT 1659 1659 E -> Q (IN REF. 3).
SQ SEQUENCE 1938 AA; 223255 MW; 387399C8F63ACCF4 CRC64;

Query Match 34.1%; Score 44; DB 1; Length 1938;
Best Local Similarity 39.1%; Pred. No. 3.5e+02;
Matches 9; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

OY 1 SDALPALADOTDALQSEEAAYVK 23
DB 1204 NDQISALTNTLDALQSKAKIEK 1226

RESULT 28
MRP4_STRPY STANDARD; PRT; 388 AA.
ID MRP4_STRPY
AC P30141;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DE Fibriinogen- and Ig-binding protein precursor (MRP protein).
GN MRP4.
OS Streptococcus pyogenes.
OC Bacteria: Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RA NCBI_TaxID=1314;
RN RN
RP STRAIN=AP4 / SEROTYPE M4;
RX MEDLINE=92269579; PubMed=1588817;
RA Stenberg L., O'Toole P., Lindahl G.;
RT "Many group A streptococcal strains express two different
RT immunoglobulin-binding proteins, encoded by closely linked genes:
RT characterization of the proteins expressed by four strains of
RT different M-type."
RL Mol. Microbiol. 6:1185-1194(1992).

```

```

RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=AP4 / SEROTYPE M4;
RX MEDLINE=92409576; PubMed=1528877;
RA O'Toole P., Stenberg L., Rissler M., Lindahl G.;
RT "Two major classes in the M protein family in group A streptococci."
RL Proc. Natl. Acad. Sci. U.S.A. 89:8661-8665(1992).
CC -1- FUNCTION: BINDS IGG MOLECULES OF THE Ig1, Ig2 AND Ig4 SUBCLASSES,
CC AND ALSO BINDS FIBRINOGEN.
CC -1- SUBCELLULAR LOCATION: TYPE I membrane protein. Cell wall.
CC -1- SIMILARITY: TO OTHER STREPTOCOCCAL AND STAPHYLOCOCCAL PROTEINS
CC IN THE REGION OF THE MEMBRANE ANCHOR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: M87831; AAA26930.1; -.
DR PIR: A46173; A46173.
DR InterPro: IPR001899; Gram_pos_anchor.
DR Pfam: PF00746; Gram_pos_anchor.1.
DR PRINTS: PR00015; GP05ANCHOR.
DR PROSITE: PS00343; GRAM_POS_ANCHORING.1.
KW IgG-binding protein; Repeat; Transmembrane; Cell wall; Signal.
FT SIGNAL 1 41
FT CHAIN 42 388 FIBRINOGEN- AND Ig-BINDING PROTEIN.
FT DOMAIN 42 363 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 364 382 MEMBRANE ANCHOR.
FT DOMAIN 383 388 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 354 359 CONSERVED IN GRAM-POSITIVE COCCI SURFACE
FT PROTEINS.
SQ SEQUENCE 388 AA; 42244 MW; 3AB19E8B8CEDA722 CRC64;

Query Match 33.7%; Score 43.5; DB 1; Length 388;
Best Local Similarity 50.0%; Pred. No. 76;
Matches 12; Conservative 5; Mismatches 6; Indels 1; Gaps 1;

OY 5 EALADOTDALQSEEAAYKADNAA 28
DB 295 KALKKQL-AKQAEELAKLKADKAS 317

RESULT 29
YTF3_RHILT STANDARD; PRT; 420 AA.
ID YTF3_RHILT
AC Q52871;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 45.4 kDa protein in tfuA 3' region (ORF3).
OS Rhizobium leguminosarum (biov. trifolii).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=386;
RA NCBI_TaxID=386;
RN RN
RP STRAIN=724;
RX MEDLINE=96312357; PubMed=8763943;
RA Breil B.T., Triplet E.W.;
RT "A newly discovered gene, tfuA, involved in the production of the
RT ribosomally synthesized peptide antibiotic trifoliotoxin."
RL J. Bacteriol. 178:4150-4156(1996).
CC -1- SIMILARITY: BELONGS TO THE UP00142 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

CC entitles requires a license agreement. (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: U39409; AAB17514.1; -  
 DR InterPro: IPR003776; DUF181.  
 DR Pfam: PF02624; DUF181.1.  
 KM Hypothetical protein.  
 SQ SEQUENCE 420 AA: 45372 MW: BEDD4ED71736238 CRC64;  
 Query Match 33.7%; Score 43.5; DB 1; Length 420;  
 Best Local Similarity 34.1%; Pred. No. 83;  
 Matches 14; Conservative 7; Mismatches 7; Indels 13; Gaps 2;  
 QY 1 SDALAL-----ADQDA-----LQSEAAVVKADNNA 28  
 Db 90 SAAMEAVETRIAEIAPADLTQATVESMRASRAMIDIDNVA 130  
 RESULT 30  
 PPB\_ECOLI STANDARD; PRT; 471 AA.  
 AC P00634; P78051; P77801;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 13-AUG-1987 (Rel. 05, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Alkaline phosphatase precursor (EC 3.1.3.1) (APASE).  
 GN PHO A B0383.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-JM83.  
 RX MEDLINE=8706741; PubMed=3537962;  
 RA Shuttleworth H., Taylor J., Minton N.;  
 RT "Sequence of the gene for alkaline phosphatase from Escherichia coli  
 JM83";  
 RL Nucleic Acids Res. 14:8689-8689(1986).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=87031576; PubMed=353724;  
 RA Chang C.N., Kuang W.-J., Chen E.Y.;  
 RT "Nucleotide sequence of the alkaline phosphatase gene of Escherichia  
 coli";  
 RL Gene 44:121-125(1986).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=88320572; PubMed=3045828;  
 RA Dubose R.F., Dykhuisen D.E., Hartl D.L.;  
 RT "Genetic exchange among natural isolates of bacteria: recombination  
 within the phoA gene of Escherichia coli";  
 RL Proc. Natl. Acad. Sci. U.S.A. 85:7036-7040(1988).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K12 / WGI555;  
 RX MEDLINE=9742617; PubMed=9278503;  
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RT "The complete genome sequence of Escherichia coli K-12";  
 RL Science 277:1453-1474(1997).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RN Duncan M., Allen E., Araujo R., Aparicio A.M., Chung E., Davis K.,  
 RA Federspiel N., Hyman R., Kaiman S., Komp C., Kudt O., Lew H.,  
 RA Lin D., Narmath A., Oetner P., Roberts D., Schramm S., Davis R.W.;  
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBD databases.  
 RN [6]  
 RP SEQUENCE OF 1-77 FROM N.A.  
 RX MEDLINE=82081850; PubMed=6273802;  
 Kikuchi Y., Yoda K., Yamasaki M., Tamura G.;

RT "The nucleotide sequence of the promoter and the amino-terminal  
 RT region of alkaline phosphatase structural gene (phoA) of Escherichia  
 RT coli";  
 RL Nucleic Acids Res. 9:5671-5678(1981).  
 RN [7]  
 RP SEQUENCE OF 23-471 (ISOZYME 3).  
 RX MEDLINE=81273081; PubMed=7022451;  
 RA Bradshaw R.A., Cancedda F., Ericsson L.H., Neumann P.A., Piccoli S.P.,  
 RA Schlesinger M.J., Shriefer K., Walsh K.A.;  
 RT "Amino acid sequence of Escherichia coli alkaline phosphatase";  
 RL Proc. Natl. Acad. Sci. U.S.A. 78:3473-3477(1981).  
 RN [8]  
 RP SEQUENCE OF 162-193 AND 385-399 FROM N.A.  
 RX MEDLINE=8211946; PubMed=7035431;  
 RA Inouye H., Barnes W., Beckwith J.;  
 RT "Signal sequence of alkaline phosphatase of Escherichia coli";  
 RL J. Bacteriol. 149:434-439(1982).  
 RN [9]  
 RP SEQUENCE OF 1-21 FROM N.A.  
 RX MEDLINE=86250586; PubMed=3522543;  
 RA Michaels S., Hunt J.F., Beckwith J.;  
 RT "Effects of signal sequence mutations on the kinetics of alkaline  
 RT phosphatase export to the periplasm in Escherichia coli";  
 RL J. Bacteriol. 167:160-167(1986).  
 RN [10]  
 RP SEQUENCE OF 1-23 FROM N.A.  
 RX MEDLINE=89340570; PubMed=2668291;  
 RA Laforet G.A., Kaiser E.T., Kendall D.A.;  
 RT "Signal peptide subsegments are not always functionally  
 RT interchangeable. M13 procoat hydrophobic core fails to transport  
 RT alkaline phosphatase in Escherichia coli";  
 RL J. Biol. Chem. 264:14478-14485(1989).  
 RN [11]  
 RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).  
 RX MEDLINE=86115281; PubMed=3910843;  
 RA Swadski J.M., Handschumacher M.D., Krishna Murthy H.M., Foster B.A.,  
 RA Wyckoff H.W.;  
 RT "Refined structure of alkaline phosphatase from Escherichia coli at  
 RT 2.8-A resolution";  
 RL J. Mol. Biol. 186:417-433(1985).  
 RN [12]  
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).  
 RX MEDLINE=91186406; PubMed=2010919;  
 RA Kim E.E., Wyckoff H.W.;  
 RT "Reaction mechanism of alkaline phosphatase based on crystal  
 RT structures. Two-metal ion catalysis";  
 RL J. Mol. Biol. 218:449-464(1991).  
 RN [13]  
 RP X-RAY CRYSTALLOGRAPHY (2.25 ANGSTROMS) OF MUTANT GLN-434.  
 RX MEDLINE=96194161; PubMed=8652582;  
 RA Ma L., Kantrowitz E.R.;  
 RT "Kinetic and X-ray structural studies of a mutant Escherichia coli  
 RT alkaline phosphatase (His-412->Gln) at one of the zinc binding  
 RT sites";  
 RL Biochemistry 35:2394-2402(1996).  
 RN [14]  
 RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).  
 RX MEDLINE=97397347; PubMed=9253408;  
 RA Murphy J.E., Stec B., Ma L., Kantrowitz E.R.;  
 RT "Trapping and visualization of a covalent enzyme-phosphate  
 RT intermediate";  
 RL Nat. Struct. Biol. 4:618-622(1997).  
 RN [15]  
 RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS).  
 RX MEDLINE=98202577; PubMed=9533886;  
 RA Stec B., Hehlir M.J., Brennan C., Nolte M., Kantrowitz E.R.;  
 RT "Kinetic and X-ray structural studies of three mutant E. coli  
 RT alkaline phosphatases: Insights into the catalytic mechanism without  
 RT the nucleophile Ser102";  
 RL J. Mol. Biol. 277:647-662(1998).  
 RN [16]  
 RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).  
 RX MEDLINE=99185045; PubMed=10085061;



RA Holtz K.M., Stec B., Kantrowitz E.R.;  
 RT "A model of the transition state in the alkaline phosphatase  
 reaction.";  
 RL J. Biol. Chem. 274:8351-8354(1999).  
 CC -1- CATALYTIC ACTIVITY: An orthophosphoric monoester + H(2)O = an  
 CC alcohol + phosphate.  
 CC -1- COFACTOR: METALLOENZYME CONTAINING TWO ZINC ATOMS AND A MAGNESIUM  
 CC ION.  
 CC -1- SUBUNIT: ISOZYMES 1 & 3 ARE A DIMER OF IDENTICAL CHAINS, ISOZYME 2  
 CC IS A DIMER OF HETEROGENEOUS CHAINS, ONE OF EACH OF THE SUBUNITS  
 CC FROM ISOZYMES 1 AND 3.  
 CC -1- SUBCELLULAR LOCATION: Periplasmic.  
 CC -1- SIMILARITY: BELONGS TO THE ALKALINE PHOSPHATASE FAMILY.  
 CC -1- DATABASE: NAME-Prozyme technical fact sheet;  
 CC WWW="http://www.prozyme.com/technical/bap.html".  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 DR EMBL: V00315; CAA23603.1; ALT\_SEQ.  
 DR EMBL: X04586; CAA28257.1; -.  
 DR EMBL: AE000145; AAC73486.1; ALT\_INT.  
 DR EMBL: U73857; AAB18107.1; -.  
 DR EMBL: M29664; AAA24365.1; -.  
 DR EMBL: M29664; AAA24364.1; -.  
 DR EMBL: M33536; AAA24372.1; -.  
 DR EMBL: M13763; AAA24358.1; -.  
 DR EMBL: J05005; AAA24362.1; -.  
 DR EMBL: J01659; AAA24359.1; -.  
 DR EMBL: J01660; AAA24360.1; -.  
 DR EMBL: J01661; AAA24361.1; -.  
 DR EMBL: J01661; CAB23260.1; -.  
 DR EMBL: J01661; CAB23261.1; -.  
 DR EMBL: M13345; AAA83893.1; -.  
 DR PIR: A00776; PAECA.  
 DR PIR: A24833; A24833.  
 DR PDB: 1AJA; 14-NOV-95.  
 DR PDB: 1AJB; 14-NOV-95.  
 DR PDB: 1AJC; 14-NOV-95.  
 DR PDB: 1AJD; 14-NOV-95.  
 DR PDB: 1ALH; 27-FEB-95.  
 DR PDB: 1ALJ; 14-NOV-95.  
 DR PDB: 1ALK; 31-JAN-94.  
 DR PDB: 2ANH; 29-JAN-96.  
 DR PDB: 1ANI; 26-JAN-96.  
 DR PDB: 1ANI; 26-JAN-96.  
 DR PDB: 1HOA; 08-MAR-96.  
 DR PDB: 1HJK; 15-OCT-97.  
 DR PDB: 1URR; 11-JUL-96.  
 DR PDB: 1URR; 11-JUL-96.  
 DR PDB: 1ELX; 27-MAY-98.  
 DR PDB: 1ELX; 27-MAY-98.  
 DR PDB: 1ELZ; 27-MAY-98.  
 DR PDB: 1B8J; 18-FEB-99.  
 DR ECODBASE: F046.6; 6TH EDITION.  
 DR Ecogene; EG10727; phoA.  
 DR InterPro: IPR001952; Alk\_\_phosphatase.  
 DR Pfam: PF00245; alk\_\_phosphatase.  
 DR PRINTS: PR00113; ALKPHPTASE.  
 DR Prodom: PD001868; Alk.phosphatase; 1.  
 DR SMART: SM00098; alkpc; 1.  
 DR PROSITE: PS00123; ALKALINE\_PHOSPHATASE; 1.  
 DR Hydrolase; Zinc; Magnesium; Phosphorylation; Periplasmic; signal;  
 KW 3D-structure; Complete proteome.  
 FT CHAIN 1 21  
 FT SIGNAL 1 21  
 FT ACT\_SITE 124 124 ALKALINE PHOSPHATASE.

FT MOD\_RES 124 124 PHOSPHORYLATION.  
 Query Match 33.7%; Score 43.5; DB 1; Length 471;  
 Best Local Similarity 31.7%; Pred. No. 93;  
 Matches 13; Conservative 6; Mismatches 9; Indels 13; Gaps 1;  
 QY 1 SDALFALADQTD-----LQSEAAVKKADNA 28  
 Db 315 NDSVPLAQMWDKAIKELSKNKGFLVQEGASIDKODHAA 355  
 RESULT 31  
 SMC3\_YEAST  
 ID SMC3\_YEAST STANDARD; PRT; 1230 AA.  
 AC P47037;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE Chromosome segregation protein SMC3 (DA-box protein SMC3).  
 GN SMC3 OR J0L074C OR J1049.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=W303;  
 RX MEDLINE=97474309; PubMed=9335333;  
 RA Michaelis C., Ciosk R., Nasmyth K.;  
 RT "Cohesins: chromosomal proteins that prevent premature separation of  
 RT sister chromatids";  
 RL Cell 91:35-45(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Rose M., Koetter P., Entian K.D.;  
 RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288C;  
 RA Sor F.J.;  
 RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: INVOLVED IN CHROMOSOME SEGREGATION IN MITOSIS. COULD BE  
 CC PART OF A CHROMOSOME CONDENSATION MOTOR.  
 CC -1- SUBCELLULAR LOCATION: Nuclear (potential).  
 CC -1- DOMAIN: CONSISTS OF TWO PUTATIVE CENTRAL COILED-COIL REGIONS  
 CC FLANKED BY PUTATIVE GLOBULAR REGIONS AT THE N- AND C-TERMINUS.  
 CC -1- SIMILARITY: BELONGS TO THE SMC FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 DR EMBL: Y14278; CAA74655.1; -.  
 DR EMBL: Z49349; CAA89366.1; -.  
 DR EMBL: X88851; CAA61313.1; -.  
 DR SGD: S0003610; SMC3.  
 DR InterPro: IPR003439; ABC\_transportr.  
 DR InterPro: IPR003405; SMC\_C.  
 DR InterPro: IPR003395; SMC\_N.  
 DR Pfam: PF02483; SMC\_C; 1.  
 DR Pfam: PF02463; SMC\_N; 1.  
 KW Mitosis; ATP-binding; Coiled coil; Nuclear protein.  
 FT NP\_BIND 32 39  
 FT DOMAIN 172 482 COILED COIL (POTENTIAL).  
 FT DOMAIN 685 1041 COILED COIL (POTENTIAL).  
 SO SEQUENCE 1230 AA; 141336 MW; B152DB8F7780341F CRC64;  
 Query Match 33.7%; Score 43.5; DB 1; Length 1230;  
 Best Local Similarity 37.0%; Pred. No. 2.5e+02;

```

Matches 10; Conservative 7; Mismatches 9; Indels 1; Gaps 1
OY 1 SDALEALADQTDALQSF-EAAYVKADN 26
    ||||| 1:1 1:1:1:1
DB 823 SDALGTTTIDSLNAELSKILPQEN 849

RESULT 32
RL4_HALHA
ID RL4_HALHA STANDARD: PRT: 248 AA.
AC Q06845.
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE 50S ribosomal protein L4E.
GN rplAe.
OS Halobacterium halobium.
OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;
OC Halobacterium.
OX NCBI_TaxID=2242;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94060115; PubMed=82421282;
RA YUKI Y., KANECHIKA R., ITOH T.:
RT "Nucleotide sequence of the genes encoding the L3, L4, and L23
RT equivalent ribosomal proteins from the archaeobacterium Halobacterium
RT halobium."
RL Biochim. Biophys. Acta 1216:335-338(1993).
CC -1- SIMILARITY: BELONGS TO THE L4E FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC DR EMBL; AB006961; BAA2271.1; -.
CC DR PIR; S43421; S43421.
CC DR InterPro; IPR002136; Ribosomal_L4/L2E.
CC DR Pfam; PF00573; Ribosomal_L4; 1.
CC DR PROSITE; PS00939; RIBOSOMAL_L2E; 1.
KW RIBOSOMAL protein.
SO SEQUENCE 248 AA; 26797 MW; 07DB50CB2E503E69 CRC64;

Query Match 33.3%; Score 43; DB 1; Length 248;
Best Local Similarity 47.6%; Pred. No. 56;
Matches 10; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

OY 4 LEALADQTDALQSEFAAYVKA 24
    ||||| 1:1:1:1:1:1
DB 158 LEALGVHADIERADEAATVRA 178

RESULT 33
TOLA_HAEIN
ID TOLA_HAEIN STANDARD: PRT: 372 AA.
AC P44678; P94810;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE TOLA protein.
GN TOLA OR H10383.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RX SPRAIN-RD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,

```

```

RA Kellavare A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKeeney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodet A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spirigs T., Hebdloom E., Cotton M.D.,
RA Uetereback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fitchman J.L., Fuhrmann J.L., Georgiades N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
RT influenzae Rd."
RL Science 269:496-512(1995).
RL [2]
RP SEQUENCE FROM N.A.
RC STRAIN-1479:
RX MEDLINE-97080550: Pubmed-8921895:
RA Sen K., Sikkema D.J., Murphy T.F.;
RT "Isolation and characterization of the Haemophilus influenzae tolQ,
RT tolR, tolA and tolB genes."
RL Gene 178:75-81(1996).
CC -1- FUNCTION: INVOLVED IN THE TONB-INDEPENDENT UPTAKE OF PROTEINS
CC (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Inner membrane
CC (Potential).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@isb-sdb.ch).
CC -----
DR EMBL: U32722: AAC22041.1;
DR EMBL: U32470: AAC44596.1;
DR HSSP: P04002: IATF.
DR TIGR: H10183.
KW Transport: Protein transport: Transmembrane: Repeat: Inner membrane:
KW Complete proteome.
FT DOMAIN 1 8 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 9 29 POTENTIAL.
FT DOMAIN 30 372 PERIPLASMIC (POTENTIAL).
FT VARIANT 48 48 V -> A (IN STRAIN 1479).
FT VARIANT 142 142 K -> R (IN STRAIN 1479).
FT VARIANT 165 165 A -> P (IN STRAIN 1479).
FT VARIANT 190 190 A -> R (IN STRAIN 1479).
FT VARIANT 203 203 V -> A (IN STRAIN 1479).
FT VARIANT 227 227 D -> A (IN STRAIN 1479).
FT VARIANT 232 232 A -> AKKAAPAKKA (IN STRAIN 1479).
FT VARIANT 234 234 T -> A (IN STRAIN 1479).
FT VARIANT 249 249 L -> F (IN STRAIN 1479).
FT VARIANT 254 254 I -> V (IN STRAIN 1479).
FT VARIANT 306 306 N -> S (IN STRAIN 1479).
FT VARIANT 323 323 T -> A (IN STRAIN 1479).
FT VARIANT 333 333 S -> P (IN STRAIN 1479).
SO SEQUENCE 372 AA; 39831 MW; 266ECP056C95544 CRC64;

Query Match 33.3%. Score 43; DB 1; Length 372;
Best Local Similarity 44.88; PRT; 587 AA;
Matches 13; Conservative 5; Mismatches 9; Indels 2; Gaps 1;

QY 2 DALE--ALADQTDALQSEAAVVKADNNA 28
:|:| :|: :|:| :|:| |
Db 130 EALEKQKQAEFEKKAKQAAEAALKTKADAEA 158

RESULT 34
ID UL25_HSVEB STANDARD: PRT; 587 AA.
AC P28928:
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 01-DEC-1992 (Rel. 24, Last annotation update)
DE Virion protein UL25.

```

GN UL25 OR 36.  
 OS Equine herpesvirus type 1 (strain Ab4p) (EHV-1).  
 OC Viruses: dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Alphaherpesvirinae; Varicelloviruses.  
 OX NCBI\_TaxID=31520;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92295566; PubMed=13186606;  
 RA Telford E.A.R., Watson M.S., McBride K., Davison A.J.;  
 RT "The DNA sequence of equine herpesvirus-1";  
 RL Virology 189:304-316(1992).  
 CC -1- FUNCTION: VIRION PROTEIN.  
 CC -1- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL25,  
 CC EHV-1 36, EBV BFRF1, HCMV UL7, ILTV ORF2, AND VZV 34.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M86664; AAB02472.1; -  
 DR PIR: B36799; M2BEC9.  
 DR InterPro: IPR002493; UL25.  
 DR Pfam: PF01499; UL25; 1.  
 SQ SEQUENCE 587 AA; 63692 MW; 6BE2F5FAB34CA3C8 CRC64;  
 QY 1 SDALBALDQTDALQSEEAAY--VKADNA 27  
 Db 98 ADALAALEDARAASADAAATPQVNASEA 126  
 QY  
 Db  
 RESULT 35  
 YQ01\_CAEEL  
 ID YQ01\_CAEEL STANDARD; PRT; 621 AA.  
 AC 009293;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE Hypothetical 69.9 kDa protein EED8.1 in chromosome II.  
 GN EED8.1.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditicoidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RA Chisoe S.;  
 RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: U23484; AAC46761.1; -  
 DR WormBep: EED8.1; CE01882.  
 DR InterPro: IPR001097; Mac\_tudor.  
 DR Pfam: PF00567; TUDOR; 1.  
 DR SMART: SM00360; RRM; 1.  
 DR PROSITE: PS50102; RRM; 1.

DR PROSITE: PS00030; RRM\_RNP\_1; FALSE\_NEG.  
 KW Hypothetical protein; RNA-binding  
 FT DOMAIN 267 346 RNA-BINDING (RRM).  
 SQ SEQUENCE 621 AA; 69894 MW; 464DB7EF91AF65C5 CRC64;  
 QY 2 DALEBALDQTDALQSEEAAYVKADNA 27  
 Db 14 DALDIISINDGVSSDEEAVSVKESS 39  
 QY  
 Db  
 RESULT 36  
 DNK\_RICPR  
 ID DNK\_RICPR STANDARD; PRT; 627 AA.  
 AC 092DX9;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Chaparone protein dnk (Heat shock protein 70) (Heat shock 70 kDa  
 DE protein) (HSP70).  
 GN DNK OR Rp185.  
 OS Rickettsia prowazekii.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 OC Rickettsiaceae; Rickettsiae; Rickettsia.  
 OX NCBI\_TaxID=782;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-MADRID E;  
 RX MEDLINE=99039489; PubMed=9823893;  
 RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,  
 RA Scharif-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,  
 RA Eriksson A.-S., Winkler H.H., Kurland C.G.;  
 RT "The genome sequence of Rickettsia prowazekii and the origin of  
 RT mitochondria".  
 RL Nature 396:133-140(1998).  
 CC -1- FUNCTION: ACTS AS A CHAPERONE (BY SIMILARITY).  
 CC -1- INDUCTION: BY STRESS CONDITIONS E.G. HEAT SHOCK (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AJ235270; CA114651.1; -  
 DR HSP; P04475; IDG4.  
 DR InterPro: IPR001023; HSP70.  
 DR Pfam: PF00012; HSP70; 1.  
 DR PRINTS: PR00301; HEATSHOCK70.  
 DR PROSITE: PS00287; HSP70\_1; 1.  
 DR PROSITE: PS00329; HSP70\_2; 1.  
 DR PROSITE: PS01036; HSP70\_3; 1.  
 KW Chaparone; ATP-binding; Heat shock; Complete proteome.  
 SQ SEQUENCE 627 AA; 68383 MW; A638A0A5A5C8AC7 CRC64;  
 QY 5 EALADQTDALQSEEAAYVK 23  
 Db 561 EALTSLKALSESADSLIK 579  
 QY  
 Db  
 RESULT 37  
 MFPL\_LYCES  
 ID MFPL\_LYCES STANDARD; PRT; 697 AA.

AC P93203;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DE 16-OCT-2001 (Rel. 40, Last annotation update)  
DR MAR binding filament-like protein 1.  
GN MPRL.  
OS Lycopersicon esculentum (Tomato).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
OC Assteridae; eunasteriids I; Solanales; Solanaceae; Solanum.  
OX NCBI\_TaxID=4081;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. VENT CHERRY; TISSUE=FRUIT;  
RX MEDLINE=97112038; PubMed=8953774;  
RA Meier I., Phelan T., Gruissem W., Spiker S., Schneider D.;  
RT "MPRL, a novel plant filament-like protein with affinity for matrix  
attachment region DNA.";  
RL Plant Cell 8:2105-2115(1996).  
CC -1- FUNCTION: BINDS DNA. INTERACTS WITH CHROMATIN VIA MATRIX  
ATTACHMENT REGIONS (MARS). LIKELY TO PARTICIPATE IN NUCLEAR  
ARCHITECTURE BY CONNECTING CHROMATIN WITH THE NUCLEAR MATRIX AND  
POTENTIALLY WITH THE NUCLEAR ENVELOPE.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR MATRIX.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL; Y07861; CAA69181.1;  
KM Nuclear protein; DNA-binding; Coiled coil.  
FT DOMAIN 125 681 COILED COIL (POTENTIAL).  
FT DOMAIN 10 15 POLY-SER.  
SQ SEQUENCE 697 AA; 79516 MW; 700A56D68D6A7E49 CRC64;  
  
Query Match 33.3%; Score 43; DB 1; Length 697;  
Best Local Similarity 43.5%; Pred. NO. 1.6e+02;  
Matches 10; Conservative 5; Mismatches 8; Indels 0; Gaps 0;  
  
OY 1 SPALNADQTDALQSEEAAYK 23  
DB 508 SDELVAASENRDLSQTELVNYK 530  
  
RESULT 38  
ID METE\_MYCTU STANDARD: PRT: 759 AA.  
AC 006584;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DE 16-OCT-2001 (Rel. 40, Last annotation update)  
DE 5-methyltetrahydropteroyltri-glutamate-homocysteine methyltransferase  
(EC 2.1.1.14) (Methionine synthase, Vitamin-B12 independent isozyme)  
DE (Cobalamin-independent methionine synthase).  
GN METE OR RV1133C OR MT1165 OR MTC2268.22.  
OS Mycobacterium tuberculosis  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteriales;  
OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1773;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=H37RV;  
RX MEDLINE=96295987; PubMed=9634230;  
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekela F.,  
RA Baddock K., Basham D., Brown D., Chillingworth T., Connor R.,  
RA Davies R., Devlin K., Fellwell T., Gentles S., Hamlin N., Holroyd S.,  
RA Hornsby T., Jagsels K., Krogh A., McLean J., Moule S., Murphy L.,  
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,

RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
RA Sulston J.E., Taylor K., Whitehead S., Barrall B.G.;  
RT "Deciphering the biology of Mycobacterium tuberculosis from the  
RT complete genome sequence.";  
RL Nature 393:537-544(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CDC 1551 / Oshkosh;  
RA Fleischman R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
RA Peterson J., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,  
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,  
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mkula A.,  
RA Bishai W.;  
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and  
RT laboratory strains.";  
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: CATALYZES THE TRANSFER OF A METHYL GROUP FROM 5-  
METHYLTETRAHYDROFOLATE TO HOMOCYSTEINE RESULTING IN METHIONINE  
FORMATION (BY SIMILARITY).  
CC -1- CATALYTIC ACTIVITY: 5-methyltetrahydropteroyl-L-glutamate + L-  
homocysteine = tetrahydropteroyl-L-glutamate + L-methionine.  
CC -1- COFACTOR: ZINC. BINDS ONE ION PER SUBUNIT (BY SIMILARITY).  
CC -1- PATHWAY: TERMINAL STEP IN THE DE NOVO BIOSYNTHESIS OF METHIONINE.  
CC -1- SIMILARITY: BELONGS TO THE VITAMIN-B12 INDEPENDENT METHIONINE  
SYNTHASE FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL; Z95585; CAB09044.1;  
DR EMBL; AE006985; AAK45422.1;  
DR TIGR; MT1165;  
DR TuberculList; RV1133C;  
DR InterPro: IPR002629; Methionine\_synth.  
DR Pfam: PF01717; Methionine\_synth; 1.  
KM Transferase: Methyltransferase; Methionine biosynthesis; Zinc; Repeat;  
KW Complete proteome.  
FT METAL 647 647 ZINC (BY SIMILARITY).  
FT METAL 649 649 ZINC (BY SIMILARITY).  
FT METAL 732 732 ZINC (BY SIMILARITY).  
SQ SEQUENCE 759 AA; 81581 MW; 702F90BF79B25C8D CRC64;  
  
Query Match 33.3%; Score 43; DB 1; Length 759;  
Best Local Similarity 43.5%; Pred. NO. 1.8e+02;  
Matches 10; Conservative 4; Mismatches 9; Indels 0; Gaps 0;  
  
OY 6 ALADQTDALQSEEAAYKADNAA 28  
DB 586 AIRDETVDLSAGIVQVDEPA 608  
  
RESULT 39  
ID YB52\_MYCPN STANDARD: PRT: 794 AA.  
AC P75034;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DE 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypothetical lipoprotein MPN152 precursor (E07\_orf794).  
GN MPN152 OR MP002.  
OS Mycoplasma pneumoniae.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;  
OC Mycoplasmataceae; Mycoplasma.  
OX NCBI\_TaxID=2104;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 29342 / M129;  
RX MEDLINE=97105885; PubMed=8948633;



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

## OM protein - nucleic search, using frame\_plus.p2n model

Run on: October 13, 2002, 04:54:53 ; Search time 8.98396 Seconds  
(without alignments)  
765.558 Million cell updates/sec

Title: US-09-847-539a-6\_COPY\_59\_86  
Perfect score: 129  
Sequence: 1 SDALFALADQTDALQSEEAAYVADNAA 28

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 38353 seqs, 122816752 residues  
Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

## Command line parameters:

-MODEL-frame.p2n.model -DB-xlh  
-O=/cg2\_1/USPTO.spool/US09847539/runat\_10102002\_093106\_5020/app\_query.fasta.1.526  
-DB-issued.patents.NA -QFMT-fastlap -SUFFIX-std.inl -MINMATCH-0.1 -LOOPEXT-0  
-LOOPEXT-0 -UNITS-bits -START-1 -END-1 -MATRIX-blosum62 -TRANS-human40.cdl  
-LIST-45 -DOCALIGN-200 -THR SCORE-pct -THR MAX-100 -THR MIN-0 -ALIGN-40  
-MODE-LOCAL -OUTFMT-pto -NORM-ext -HEAPSIZE-500 -MINLEN-0 -MAXLEN-2000000000  
-USRR-US09847539.ecgn\_1.1.13-runat\_10102002\_093106\_5020 -NCPU-6 -TCP-3  
-NO\_XLPRY -NO\_MMAP -LARGEQUERY -NEG\_SCORES-0 -WAIT -LONGLOG -DEV\_TIMEOUT-120  
-MARN\_TIMEOUT-30 -THREADS-1 -XGAPOP-10 -XGAPEXT-0.5 -FGAPOP-6 -FGAPEXT-7  
-YGAPOP-10 -YGAPEXT-0.5 -DELOP-6 -DELEXT-7

## Database :

Issued.Patents.NA.\*  
1: /cg2\_6/ptodata/2/ina/5A.COMB.seq.\*  
2: /cg2\_6/ptodata/2/ina/5B.COMB.seq.\*  
3: /cg2\_6/ptodata/2/ina/6A.COMB.seq.\*  
4: /cg2\_6/ptodata/2/ina/6B.COMB.seq.\*  
5: /cg2\_6/ptodata/2/ina/PCPUS.COMB.seq.\*  
6: /cg2\_6/ptodata/2/ina/backfile1.seq.\*

## SUMMARIES

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	DB ID	Description
1	47	36.4	59065	4	US-09-813-817-3
2	47	36.4	4403765	4	US-09-103-840A-2
3	47	36.4	4411529	4	US-09-103-840A-1
4	45	34.9	4403765	4	US-09-103-840A-1
5	45	34.9	4411529	4	US-09-103-840A-2
6	44.5	34.5	1548	2	US-08-762-106-5
7	44.5	34.5	1548	4	US-09-320-774-5
8	44.5	34.5	1581	2	US-08-762-106-6
9	44.5	34.5	1581	2	US-08-762-106-6
10	44	34.1	1033	3	US-09-320-774-6
11	44	34.1	1772	2	US-08-961-083-191
12	43.5	33.7	744	2	US-08-960-022-13
					Sequence 19, Appl

C	13	43.5	33.7	744	3	US-08-526-840B-19	Sequence 19, Appl
	14	43.5	33.7	1149	1	US-08-049-282B-1	Sequence 1, Appl1
	15	43.5	33.7	1149	2	US-08-537-715-1	Sequence 1, Appl1
	16	43.5	33.7	1149	5	PCT-US94-04173-1	Sequence 1, Appl1
	17	43.5	33.7	1454	2	US-08-657-392-1	Sequence 1, Appl1
	18	43.5	33.7	1454	5	PCT-US94-02539-1	Sequence 1, Appl1
	19	43.5	33.7	1455	2	US-08-657-392-26	Sequence 26, Appl
	20	43.5	33.7	1455	5	PCT-US94-02539-26	Sequence 26, Appl
	21	43.5	33.7	1857	5	US-08-537-715-3	Sequence 3, Appl1
	22	43.5	33.7	1857	5	PCT-US94-04173-3	Sequence 3, Appl1
	23	43.5	33.7	2233	1	US-08-235-838-6	Sequence 6, Appl1
	24	43.5	33.7	2233	4	US-08-465-473B-6	Sequence 6, Appl1
	25	43.5	33.7	2681	2	US-08-928-213B-7	Sequence 7, Appl1
	26	43.5	33.7	6162	2	US-08-673-312-9	Sequence 9, Appl1
	27	43	33.3	1221	2	US-08-934-846-3	Sequence 3, Appl1
	28	43	33.3	1221	4	US-08-238-557-3	Sequence 3, Appl1
	29	43	33.3	1254	2	US-08-934-846-1	Sequence 1, Appl1
	30	43	33.3	1254	4	US-08-238-557-1	Sequence 1, Appl1
	31	43	33.3	4224	1	US-08-612-521-1	Sequence 1, Appl1
	32	42.5	32.9	1274	2	US-08-470-670A-1	Sequence 1, Appl1
	33	42.5	32.9	1274	4	US-08-461-511A-1	Sequence 1, Appl1
	34	42.5	32.9	1274	5	PCT-US94-07089-1	Sequence 1, Appl1
	35	42.5	32.9	1303	1	US-08-081-448-1	Sequence 1, Appl1
	36	42.5	32.9	43280	2	US-08-804-227C-1	Sequence 1, Appl1
	37	42	32.6	621	2	US-08-943-915-4	Sequence 1, Appl1
	38	42	32.6	927	4	US-09-254-733-4	Sequence 4, Appl1
	39	42	32.6	1060	1	US-08-090-013-1	Sequence 1, Appl1
	40	42	32.6	1060	1	US-08-081-328-1	Sequence 1, Appl1
	41	42	32.6	1060	1	US-08-232-249-1	Sequence 1, Appl1
	42	42	32.6	1060	2	US-08-921-426-7	Sequence 7, Appl1
	43	42	32.6	1060	2	US-08-833-642A-1	Sequence 1, Appl1
	44	42	32.6	1060	2	US-08-140-008A-3	Sequence 3, Appl1
	45	42	32.6	1060	2	US-08-389-423-1	Sequence 1, Appl1

## ALIGNMENTS

RESULT 1  
US-09-813-817-3/c  
: Sequence 3, Application US/09813817  
: Patent No. 6340583  
: GENERAL INFORMATION:  
: APPLICANT: YAN, Chunhua et al.  
: TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC  
: TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES  
: FILE REFERENCE: CLO01178  
: CURRENT APPLICATION NUMBER: US/09/813, 817  
: CURRENT FILING DATE: 2001-03-22  
: NUMBER OF SEQ ID NOS: 4  
: SOFTWARE: FastSeq for Windows Version 4.0  
: SEQ ID NO 3  
: LENGTH: 59065  
: TYPE: DNA  
: ORGANISM: Human  
US-09-813-817-3

Alignment Scores:  
Pred. No.: 2.7e+03  
Score: 47.00  
Percent Similarity: 76.47%  
Best Local Similarity: 58.82%  
Query Match: 36.43%  
DB: 4  
Gaps: 0

US-09-847-539a-6\_COPY\_59\_86 (1-28) x US-09-813-817-3 (1-59065)  
Qy 8 AlaasgintrrasppalaaleuglunlaiaaValvalysala 24  
Db 21239 GCTGAACCAACTGACTCTTAGAACCCAGCTGCTTCATAAGCG 21189  
RESULT 2  
US-09-103-840A-2/c

```

? Sequence 2 Application US/09103840A
? Patent No. 6294328
? GENERAL INFORMATION:
? APPLICANT: FLEISCHMAN, Robert D.
? APPLICANT: WHITE, Owen R.
? APPLICANT: FRASER, Claire M.
? APPLICANT: VENTER, John C.
? TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
? TITLE OF INVENTION: TUBERCULOSIS
? FILE REFERENCE: 24366-20007.00
? CURRENT APPLICATION NUMBER: US/09/103,840A
? CURRENT FILING DATE: 1998-06-24
? NUMBER OF SEQ ID NOS: 2
? SOFTWARE: Patentin Ver. 2.1
? SEQ ID NO 2
? LENGTH: 4403765
? TYPE: DNA
? ORGANISM: Mycobacterium tuberculosis
? FEATURE:
? OTHER INFORMATION: CDC 1551
? OTHER INFORMATION: "n" bases at various positions throughout the sequence
? US-09-103-840A-2

```

Alignment Scores:	
Pred. No.:	4,15e+05
Score:	47.00
Percent Similarity:	54.55%
Best Local Similarity:	40.91%
Query Match:	36.43%
DB:	4
	Gaps: 0

Oy 7 leuAlaaspGlnThrAspAlaIeuGluSerGluGluAlaValValLysAlaAspSn 26  
DB 4036256 TTGATGACGCCCGCCAGCCTGTGGAAAGGCAGCAAGCTAACCATTCGTGATTTCGACAC 4036197

QY 27 ALALA 28  
|||  
Db 4036196 GGTGCT 4036191

```

RESULT 3
US-09-103-840A-1/C
: Sequence 1, Application US/09103840A
: Patent No. 6294328
:
: GENERAL INFORMATION:
:
: APPLICANT: FLEISCHMAN, Robert D.
: APPLICANT: WHITE, Owen R.
: APPLICANT: FRASER, Claire M.
: APPLICANT: VENTER, John C.
:
: TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
: TUBERCULOSIS
:
: FILE REFERENCE: 24366-20007.00
:
: CURRENT APPLICATION NUMBER: US/09/103,840A
:
: CURRENT FILING DATE: 1998-06-24
:
: NUMBER OF SEQ ID NOS: 2
:
: SOFTWARE: Patentln Ver. 2.1
:
: SEQ ID NO 1
:
: LENGTH: 4411529
:
: TYPE: DNA
:
: ORGANISM: Mycobacterium tuberculosis
:
: OTHER INFORMATION: H37Rv
:
: US-09-103-840A-1

```

Prediction Scores:	
Pred. NO :	4.16e+05
Score:	47.00
Percent Similarity:	54.55%
Best local Similarity:	40.91%
Query Match:	36.43%
DB:	
	4
Caps:	0
	4411529
Length:	9
Matches:	3
Conservative:	9
Mismatches:	10
Indels:	0
Gaps:	0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-09-103-840A-1 (1-4411529)

QY 7 leuAlaaspGrInThAspaAlaLeuInserGluGluAlaAlaValValIlysAlaAspAsn 26  
|||||  
Db 4044185 TTGATGACAGCCGCCGACCTGGTGGAGGCGACCAAGTACCATCCGCGATATCGACAAAC 40441286

QY 27 ALA1A 28  
||  
Db 4044125 GGTGCT 4044120

RESULT 4  
US-09-103-840A-2

SEQUENCE 2, APPLICATION NO. 6294328

APPLICANT: FLEISCHMAN, Robert D.  
APPLICANT: WHITE, Owen B.

APPLICANT: FRASER, Claire M. -  
APPLICANT: VENTER, John C.

; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
; TUBERCULOSIS  
; TITLE OF INVENTION: TUBERCULOSIS

```

; FILE REFERENCE: 24366-2000/.00
; CURRENT APPLICATION NUMBER: US/09/103,840A

```

NUMBER OF SEQ ID NOS: 2

; SEQ ID NO 2  
LENGTH: 4403765

```

; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis

```

```

;
; FEATURE:
; OTHER INFORMATION: CDC 1551
;

```

```

; OTHER INFORMATION: "n" bases at various positions throughout the sequence
;
; OTHER INFORMATION: represent a, t, c or g
;

```

US-09-103-840A-2

Pred. No.:	6.3e+05	Length:	4403765
Score:	45.00	Matches:	13
Percent Similarity:	55.88%	Conservative:	6
Best Local Similarity:	38.2%	Mismatches:	7
Query Match:	34.88%	Indels:	8
DB:	4	Gaps:	1

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-09-103-840A-2 (1-4403765)

QY 3 ALaLeuGluaLaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluaAlaAlaValVal 22

Db 1667287 GCCAACCGAGCGCCCTGCAAGACCTGAGCGACGAGGTTTCAGGCCGAACAGGAAGCCTTAAC 1667346

```

QY      23  LysAla-----AspAsnAlaAla 28
          |||||          |||||
DB 1667347 AAGCGATGTCGACGTGAAACCGCTCGGGACACGCTGCC 1667388

```

RESULT 5  
US-09-103-840A-1

Patent No. 6294328

APPLICANT: FLEISCHMAN, Robert D.  
APPLICANT: WHITE, Owen R.

APPLICANT: FRASER, Claire M.  
APPLICANT: VENTER, John C.

; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
; TITLE OF INVENTION: TUBERCULOSIS

FILE REFERENCE: 24508-20007.00  
CURRENT APPLICATION NUMBER: US/09/103,840A  
CANDIDATE BIRTH DATE: 1900-05-24

CONTRACT ID NO: 1000  
NUMBER OF SEQ ID NOS: 2  
SOFTWARE: DnaStar ver 3.1

```

; SEQ ID NO 1
; LENGTH: 4411529

```

```

; LENGTH: 4411529
; TYPE: DNA

```



ORGANISM: Mycobacterium tuberculosis  
OTHER INFORMATION: H37RV  
US-09-103-840A-1

## Alignment Scores:

Pred. No.:	6.3e+05	Length:	4411529
Score:	45.00	Matches:	13
Percent Similarity:	55.88%	Conservative:	6
Best Local Similarity:	38.24%	Mismatches:	7
Query Match:	34.88%	Indels:	8
DB:	4	Gaps:	1

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-09-103-840A-1 (1-4411529)

OY 3 ALaLeuGluaLaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaVal 22

DB 1667153 GCCAACGACGCCCTGCAAGACCTGACGAGCTTCAGCGCGAAGAAAGGCTTAAC 1667212

OY 23 LysAla-----AspAsnAlaAla 28

DB 1667213 AAGCGCATGTCGACGTGGAAACCGCTCGGACAAACGCTGCC 1667254

## RESULT 6

US-08-762-106-5  
Sequence 5, Application US/08762106  
Patent No. 5948677

## GENERAL INFORMATION:

APPLICANT: Jarvik, Jonathan W.  
TITLE OF INVENTION: READING FRAME INDEPENDENT EPIOTOPE  
TAGGING  
NUMBER OF SEQUENCES: 47  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Harris Brotman  
STREET: 202 Coast Blvd., Suite 111  
CITY: La Jolla  
STATE: California

COUNTRY: US  
ZIP: 92037

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/762,106  
FILING DATE: 09-DEC-1996

## CLASSIFICATION: 435

## ATTORNEY/AGENT INFORMATION:

NAME: Brotman, Harris F.  
REGISTRATION NUMBER: 35,461

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 654-2428

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 1548 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

HYPOTHETICAL: NO

ANTI-SENSE: NO

US-08-762-106-5

## Alignment Scores:

Pred. No.:	76.7	Length:	1548
Score:	44.50	Matches:	15
Percent Similarity:	65.52%	Conservative:	4
Best Local Similarity:	51.72%	Mismatches:	7
Query Match:	34.50%	Indels:	3
DB:	2	Gaps:	2

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-08-762-106-5 (1-1548)

OY 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAla 20

DB 1105 GCCGATGCCGTGAGCGCGCGCGGAGCTGACGGCG---CAGCGGAGGAGCGCGCC 1161

OY 21 ValValLys-----AlaAspAsnAla 27

DB 1162 AACGCCAAGTGGAGCGGACAAAGCGC 1188

## RESULT 7

US-09-320-774-5  
Sequence 5, Application US/09320774  
Patent No. 6265545

## GENERAL INFORMATION:

APPLICANT: Jarvik, Jonathan W.  
TITLE OF INVENTION: READING FRAME INDEPENDENT EPIOTOPE  
TAGGING  
NUMBER OF SEQUENCES: 47  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Harris Brotman  
STREET: 202 Coast Blvd., Suite 111  
CITY: La Jolla  
STATE: California

COUNTRY: US  
ZIP: 92037

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/320,774  
FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/762,106  
FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Brotman, Harris F.  
REGISTRATION NUMBER: 35,461

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 654-2428

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 1548 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

HYPOTHETICAL: NO

ANTI-SENSE: NO

US-09-320-774-5

## Alignment Scores:

Pred. No.:	76.7	Length:	1548
Score:	44.50	Matches:	15
Percent Similarity:	65.52%	Conservative:	4
Best Local Similarity:	51.72%	Mismatches:	7
Query Match:	34.50%	Indels:	3
DB:	4	Gaps:	2

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-09-320-774-5 (1-1548)

OY 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAla 20

DB 1105 GCCGATGCCGTGAGCGCGCGCGGAGCTGACGGCG---CAGCGGAGGAGCGCGCC 1161

OY 21 ValValLys-----AlaAspAsnAla 27

DB 1162 AACGCCAAGTGGAGCGGACAAAGCGC 1188

## RESULT 8

US-08-762-106-6  
Sequence 6, Application US/08762106

```
? Patent No. 5948677
? GENERAL INFORMATION:
? APPLICANT: Jarvik, Jonathan W.
? TITLE OF INVENTION: READING FRAME INDEPENDENT EPITOPE
? TITLE OF INVENTION: TAGGING
? NUMBER OF SEQUENCES: 47
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Harris Brotman
? STREET: 202 Coast Blvd., Suite 111
? CITY: La Jolla
? STATE: California
? COUNTRY: US
? ZIP: 92037
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patentin Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/762,106
? FILING DATE: 09-DEC-1996
? CLASSIFICATION: 435
? ATTORNEY/AGENT INFORMATION:
? NAME: Brotman, Harris F.
? REGISTRATION NUMBER: 35,461
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (619) 654-2428
? INFORMATION FOR SEQ ID NO: 6:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1581 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: DNA (genomic)
? HYPOTHEICAL: NO
? ANTI-SENSE: NO
? US-08-762-106-6

Alignment Scores:
Pred. No.: 78.7 Length: 1581
Score: 44.50 Matches: 15
Percent Similarity: 65.52% Conservative: 4
Best Local Similarity: 51.72% Mismatches: 7
Query Match: 34.50% Indels: 3
DB: 2 Gaps: 2

US-09-847-539a-6_COPY_59_86 (1-28) x US-08-762-106-6 (1-1581)
QY 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAla 20
Db 1138 GCCGATCCCTCGACGCGCGCGCGGCGGAGCTGACGCGC---CAGCGCGAGAGCGCGC 1194
QY 21 ValValIys-----AlaAspAsnAla 27
Db 1195 AACGCCAAGTGGAGCGCGGACAAAGGCG 1221

RESULT 9
US-09-320-774-6
? Sequence 6, Application US/09320774
? Patent No. 6265545
? GENERAL INFORMATION:
? APPLICANT: Jarvik, Jonathan W.
? TITLE OF INVENTION: READING FRAME INDEPENDENT EPITOPE
? TITLE OF INVENTION: TAGGING
? NUMBER OF SEQUENCES: 47
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Harris Brotman
? STREET: 202 Coast Blvd., Suite 111
? CITY: La Jolla
? STATE: California
? COUNTRY: US
? ZIP: 92037
? COMPUTER READABLE FORM:
```

```
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patentin Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/09/320,774
? FILING DATE:
? CLASSIFICATION:
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/762,106
? FILING DATE:
? ATTORNEY/AGENT INFORMATION:
? NAME: Brotman, Harris F.
? REGISTRATION NUMBER: 35,461
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (619) 654-2428
? INFORMATION FOR SEQ ID NO: 6:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1581 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: DNA (genomic)
? HYPOTHEICAL: NO
? ANTI-SENSE: NO
? US-09-320-774-6

Alignment Scores:
Pred. No.: 78.7 Length: 1581
Score: 44.50 Matches: 15
Percent Similarity: 65.52% Conservative: 4
Best Local Similarity: 51.72% Mismatches: 7
Query Match: 34.50% Indels: 3
DB: 2 Gaps: 2

US-09-847-539a-6_COPY_59_86 (1-28) x US-09-320-774-6 (1-1581)
QY 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAla 20
Db 1138 GCCGATCCCTCGACGCGCGCGGCGGAGCTGACGCGC---CAGCGCGAGAGCGCGC 1194
QY 21 ValValIys-----AlaAspAsnAla 27
Db 1195 AACGCCAAGTGGAGCGCGGACAAAGGCG 1221

RESULT 10
US-09-961-083-191
? Sequence 191, Application US/08961083
? Patent No. 6159469
? GENERAL INFORMATION:
? APPLICANT: Chol et. al.
? TITLE OF INVENTION: Streptococcus pneumoniae Antigenes and Vaccines
? NUMBER OF SEQUENCES: 452
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Human Genome Sciences, Inc.
? STREET: 9410 Key West Avenue
? CITY: Rockville
? STATE: Maryland
? COUNTRY: USA
? ZIP: 20850
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
? COMPUTER: HP Vectra 486/33
? OPERATING SYSTEM: MSDOS version 6.2
? SOFTWARE: ASCII Text
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/961,083
? FILING DATE:
? CLASSIFICATION: 435
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER:
? FILING DATE:
? ATTORNEY/AGENT INFORMATION:
```

NAME: Brookes, A. Anders  
REGISTRATION NUMBER: 36,373  
REFERENCE/DOCKET NUMBER: PB340P2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 191:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1033 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-961-083-191

Alignment Scores:  
Pred. No.: 56.4 Length: 1033  
Score: 44.00 Matches: 11  
Percent Similarity: 65.00% Conservative: 2  
Best local Similarity: 55.00% Mismatches: 7  
Query Match: 34.11% Indels: 0  
DB: 3 Gaps: 0

US-09-847-539a-6\_COPY\_59\_86 (1-28) x US-08-961-083-191 (1-1033)

QY 8 AlaaspGlnThraspAlaLeuGlnSerGluGlnAlaValLysAlaAspAsnAla 27  
DB 8 GCAGGCGAGCAGATGCTCCGCAATTGAAAGCGCGAGTACCGAGGAGAAAAACA 67

RESULT 11  
US-08-960-022-13  
Sequence 13, Application US/08960022  
Patent No. 5976837  
GENERAL INFORMATION:  
APPLICANT: Jacobs, Kenneth  
APPLICANT: McCoy, John M.  
APPLICANT: Lavallee, Edward R.  
APPLICANT: Racie, Lisa A.  
APPLICANT: Merberg, David  
APPLICANT: Treacy, Maurice  
APPLICANT: Spaulding, Vikki  
APPLICANT: Agostino, Michael J.  
TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES  
TITLE OF INVENTION: ENCODING THEM  
NUMBER OF SEQUENCES: 30  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genetics Institute, Inc.  
STREET: 87 Cambridgepark Drive  
CITY: Cambridge  
STATE: MA  
COUNTRY: U.S.A.  
ZIP: 02140  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/960,022  
FILING DATE:  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Sprunger, Suzanne A.  
REGISTRATION NUMBER: 41,323  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 498-8284  
TELEFAX: (617) 876-5851  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1772 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA

US-08-960-022-13

Alignment Scores:  
Pred. No.: 111 Length: 1772  
Score: 44.00 Matches: 9  
Percent Similarity: 68.00% Conservative: 8  
Best local Similarity: 36.00% Mismatches: 8  
Query Match: 34.11% Indels: 0  
DB: 2 Gaps: 0

US-09-847-539a-6\_COPY\_59\_86 (1-28) x US-08-960-022-13 (1-1772)

QY 3 AlaLeuGlnAlaLeuAlaAspGlnThraspAlaLeuGlnSerGluGlnAlaValLysAla 22  
DB 795 AGCATCAAGCGCAAGAGCGAGCAGCAGCTGCTGAAGAAAGAGAGCGACGCTGTG 854

QY 23 LysAlaAspAsnAla 27  
DB 855 GAAAGTAACAACGCA 869

RESULT 12  
US-08-743-637B-19/C  
Sequence 19, Application US/08743637B  
Patent No. 5994066  
GENERAL INFORMATION:  
APPLICANT: BERGERON, Michel G.  
APPLICANT: PICARD, Francois J.  
APPLICANT: OUELLETTE, Marc  
APPLICANT: ROY, Paul H.  
TITLE OF INVENTION: SPECIES-SPECIFIC AND UNIVERSAL DNA  
TITLE OF INVENTION: PROBES AND AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND  
IDENTIFY COMMON BACTERIAL PATHOGENS AND ASSOCIATED  
TITLE OF INVENTION: ANTIBIOTIC RESISTANCE GENES FROM CLINICAL SPECIMENS ...  
NUMBER OF SEQUENCES: 273  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: OVARLES & BRADY  
STREET: 411 EAST WISCONSIN AVENUE  
CITY: MILWAUKEE  
STATE: WISCONSIN  
COUNTRY: USA  
ZIP: 53202-4497  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/743,637B  
FILING DATE: 04-NOV-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/526,840  
FILING DATE: 11-SEP-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: BAKER, Jean C.  
REGISTRATION NUMBER: 35,433  
REFERENCE/DOCKET NUMBER: 850586,90012  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (414) 277-5000  
TELEFAX: (414) 277-5591  
INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 744 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: Pseudomonas aeruginosa  
US-08-743-637B-19

Alignment Scores:  
Pred. No.: 45.7 Length: 744



OY 23 Lys 23  
|||  
Db 787 AAG 789

## RESULT 15

US-08-537-715-1  
; Sequence 1, Application US/08537715  
; Patent No. 5910627  
; GENERAL INFORMATION:  
; APPLICANT: Chuck, George S.  
; APPLICANT: Dooner, Hugo K.  
; APPLICANT: Courtney-Guterson, Neal  
; APPLICANT: Keller, Janis  
; APPLICANT: Nijjar, Charanjit S.  
; APPLICANT: Ralston, Edward J.  
; TITLE OF INVENTION: PH Genes and Their Uses  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Center, Eighth Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-3834  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/537,715  
; FILING DATE: 16-FEB-1996  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/049,282  
; FILING DATE: 16-APR-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: WO PCT/US94/04173  
; FILING DATE: 15-APR-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Bastian, Kevin L.  
; REGISTRATION NUMBER: 34,774  
; REFERENCE/DOCKET NUMBER: 012176-003410US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 576-0200  
; TELEFAX: (415) 576-0300  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1149 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: CDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..1149  
; OTHER INFORMATION: /note="Ph gene cDNA"  
US-08-537-715-1

## Alignment Scores:

Pred. No.:	78.8	Length:	1149
Score:	43.50	Matches:	11
Percent Similarity:	76.198	Conservative:	5
Best Local Similarity:	52.388	Mismatches:	4
Query Match:	33.724	Indels:	1
DB:	2	Gaps:	1

US-09-847-539a-6\_COPY\_59\_86 (1-28) x US-08-537-715-1 (1-1149)

OY 4 LeuGluAlaLeuAlaAspGlnThrAspAla---LeuGlnSerGluGluAlaAlaVal 22  
||||||| |||:|||||:||||| |||||:|||||:||||| |||  
Db 727 CTTGAAGCTAGAGCAATCATCAGACGAGGCTACGCTGCAGACAAAGATTACAGGTACTGTG 786

OY 23 Lys 23  
|||  
Db 787 AAG 789

## RESULT 16

PCT-US94-04173-1  
; Sequence 1, Application PC/TUS9404173  
; GENERAL INFORMATION:  
; APPLICANT:  
; TITLE OF INVENTION: PH GENES AND THEIR USES  
; NUMBER OF SEQUENCES: 4  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US94/04173  
; FILING DATE: 15-APR-1994  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/049,282  
; FILING DATE: 16-APR-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Bastian, Kevin L.  
; REGISTRATION NUMBER: 34,774  
; REFERENCE/DOCKET NUMBER: 12176-34-1PC  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 543-9600  
; TELEFAX: (415) 543-5043  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1149 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: CDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..1146  
; PCT-US94-04173-1

## Alignment Scores:

Pred. No.:	78.8	Length:	1149
Score:	43.50	Matches:	11
Percent Similarity:	76.198	Conservative:	5
Best Local Similarity:	52.388	Mismatches:	4
Query Match:	33.724	Indels:	1
DB:	5	Gaps:	1

US-09-847-539a-6\_COPY\_59\_86 (1-28) x PCT-US94-04173-1 (1-1149)

OY 4 LeuGluAlaLeuAlaAspGlnThrAspAla---LeuGlnSerGluGluAlaAlaVal 22  
||||||| |||:|||||:||||| |||||:|||||:||||| |||  
Db 727 CTTGAAGCTAGAGCAATCATCAGACGAGGCTACGCTGCAGACAAAGATTACAGGTACTGTG 786

OY 23 Lys 23  
|||  
Db 787 AAG 789

RESULT 17  
US-08-657-392-1  
; Sequence 1, Application US/08657392  
; Patent No. 5843634  
; GENERAL INFORMATION:  
; APPLICANT: Biate, E.M.  
; APPLICANT: Brennan, C.A.  
; APPLICANT: Bridon, D.P.  
; APPLICANT: Jaffe, K.D.  
; APPLICANT: Krafft, G.A.  
; APPLICANT: Mandelk, W.  
; APPLICANT: March, S.C.  
; APPLICANT: Russell, J.R.

```

: APPLICANT: Yue, V.T.
: TITLE OF INVENTION: Genetically Engineered Enzymes And Their
: TITLE OF INVENTION: Conjugates For Diagnostic Assays
: NUMBER OF SEQUENCES: 34
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: ABBOTT LABORATORIES
: STREET: One Abbott Park Road
: CITY: Abbott Park
: STATE: Illinois
: COUNTRY: USA
: ZIP: 60064-3500
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: SoftPC
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/657,392
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/100,708
: FILING DATE: July 29, 1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Wong, Wean Khing
: REGISTRATION NUMBER: 33,561
: REFERENCE/DOCKET NUMBER: 5324.US.P1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (708) 938-3517
: TELEFAX: (708) 938-2623
: TELEX:
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1454 nucleotides
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE:
: ORIGINAL SOURCE:
: ORGANISM: Escherichia coli
: US-08-657-392-1

Alignment Scores:
Pred. No.: 106 Length: 1454
Score: 43.50 Matches: 13
Percent Similarity: 46.34% Conservative: 6
Best Local Similarity: 31.71% Mismatches: 9
Query Match: 33.72% Indels: 13
DB: Gaps: 1

US-09-847-539a-6_COPY_59_86 (1-28) x US-08-657-392-1 (1-1454)
OY 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAla----- 13
DB 968 AACGACCTGGTTCACACCTGGCGCAGATGACCGCAAAAGCCATTGAACCTTTAAGTAAA 1027
OY 14 -----LeuGlnSerGluGluAlaAlaValAlaValAlaAspAsnAla 27
DB 1028 AATGAGAAAGCTTTTCTCTGCAAGTTGAAGTGGCTCATGATGATAAACAGATCATGCT 1087
OY 28 Ala 28
DB 1088 GCG 1090

RESULT 18
PCT-US94-02539-1
: Sequence 1, Application PC/TUS9402539
: GENERAL INFORMATION:
: APPLICANT: Brate, E.M.
: APPLICANT: Brennan, C.A.
: APPLICANT: Bridon, D.P.
: APPLICANT: Jaffe, K.D.
: APPLICANT: Kraft, G.A.
```

```

: APPLICANT: Mandeckl, W.
: APPLICANT: March, S.C.
: APPLICANT: Russell, J.R.
: APPLICANT: Yue, V.T.
: TITLE OF INVENTION: Genetically Engineered Enzymes
: TITLE OF INVENTION: And Their
: TITLE OF INVENTION: Conjugates For Diagnostic Assays
: NUMBER OF SEQUENCES: 34
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: ABBOTT LABORATORIES
: STREET: One Abbott Park Road
: CITY: Abbott Park
: STATE: Illinois
: COUNTRY: USA
: ZIP: 60064-3500
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: SoftPC
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US94/02539
: FILING DATE:
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: Wong, Wean Khing
: REGISTRATION NUMBER: 33,561
: REFERENCE/DOCKET NUMBER: 5324.PC.01
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (708) 938-3517
: TELEFAX: (708) 938-2623
: TELEX:
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1454 nucleotides
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE:
: ORIGINAL SOURCE:
: ORGANISM: Escherichia coli
: PCT-US94-02539-1

Alignment Scores:
Pred. No.: 106 Length: 1454
Score: 43.50 Matches: 13
Percent Similarity: 46.34% Conservative: 6
Best Local Similarity: 31.71% Mismatches: 9
Query Match: 33.72% Indels: 13
DB: Gaps: 1

US-09-847-539a-6_COPY_59_86 (1-28) x PCT-US94-02539-1 (1-1454)
OY 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAla----- 13
DB 968 AACGACCTGGTTCACACCTGGCGCAGATGACCGCAAAAGCCATTGAACCTTTAAGTAAA 1027
OY 14 -----LeuGlnSerGluGluAlaAlaValAlaValAlaAspAsnAla 27
DB 1028 AATGAGAAAGCTTTTCTCTGCAAGTTGAAGTGGCTCATGATGATAAACAGATCATGCT 1087
OY 28 Ala 28
DB 1088 GCG 1090

RESULT 19
US-08-657-392-26
: Sequence 26, Application US/08657392
: Patient No. 5843634
: GENERAL INFORMATION:
: APPLICANT: Brate, E.M.
: APPLICANT: Brennan, C.A.
: APPLICANT: Bridon, D.P.
```



```

: Patent No. 5910627
: GENERAL INFORMATION:
: APPLICANT: Chuck, George S.
: APPLICANT: Dooner, Hugo K.
: APPLICANT: Courtney-Guterson, Neal
: APPLICANT: Keller, Janis
: APPLICANT: Nijjar, Charnjit S.
: APPLICANT: Ralston, Edward J.
: TITLE OF INVENTION: PH Genes and Their Uses
: NUMBER OF SEQUENCES: 4
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Townsend and Townsend and Crew LLP
: STREET: Two Embarcadero Center, Eighth Floor
: CITY: San Francisco
: STATE: California
: COUNTRY: USA
: ZIP: 94111-3834
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/537,715
: FILING DATE: 16-FEB-1996
: CLASSIFICATION: 800
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/049,282
: FILING DATE: 16-APR-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: WO PCT/US94/04173
: FILING DATE: 15-APR-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Bastian, Kevin L.
: REGISTRATION NUMBER: 34,774
: REFERENCE/DOCKET NUMBER: 012176-003410US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 576-0200
: TELEFAX: (415) 576-0300
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1857 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 481..1632
: OTHER INFORMATION: /note="ph6 gene cDNA"
: US-09-537-715-3

Alignment Scores:
Pred. No.: 144 Length: 1857
Score: 43.50 Matches: 11
Percent Similarity: 76.19% Conservative: 5
Best Local Similarity: 52.38% Mismatches: 4
Query Match: 33.72% Indels: 1
DB: 2 Gaps: 1

US-09-847-539a-6_COPY_59_86 (1-28) x US-08-537-715-3 (1-1857)
Qy 4 LeuGluAlaLeuAlaSpGlnThraSPAla---LeuGlnSerGluGluAlaValVal 22
Db 1210 CTTGAAGCTAGAGCCATCAGACGAGGCTACGCTCCAGACAAAGATACAGGTA 1269
Qy 23 Lys 23
Db 1270 AAG 1272

RESULT 22
PCT-US94-04173-3
: Sequence 3, Application PC/TUS9404173
```

```

: GENERAL INFORMATION:
: APPLICANT:
: TITLE OF INVENTION: PH GENES AND THEIR USES
: NUMBER OF SEQUENCES: 4
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US94/04173
: FILING DATE: 15-APR-1994
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/049,282
: FILING DATE: 16-APR-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Bastian, Kevin L.
: REGISTRATION NUMBER: 34,774
: REFERENCE/DOCKET NUMBER: 12176-34-1PC
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 543-5043
: TELEFAX: (415) 543-5043
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1857 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 481..1632
: PCT-US94-04173-3

Alignment Scores:
Pred. No.: 144 Length: 1857
Score: 43.50 Matches: 11
Percent Similarity: 76.19% Conservative: 5
Best Local Similarity: 52.38% Mismatches: 4
Query Match: 33.72% Indels: 1
DB: 5 Gaps: 1

US-09-847-539a-6_COPY_59_86 (1-28) x PCT-US94-04173-3 (1-1857)
Qy 4 LeuGluAlaLeuAlaSpGlnThraSPAla---LeuGlnSerGluGluAlaValVal 22
Db 1210 CTTGAAGCTAGAGCCATCAGACGAGGCTACGCTCCAGACAAAGATACAGGTA 1269
Qy 23 Lys 23
Db 1270 AAG 1272

RESULT 23
US-08-235-838-6
: Sequence 6, Application US/08235838
: Patent No. 5571894
: GENERAL INFORMATION:
: APPLICANT: Wels, Winfried S.
: APPLICANT: Hynes, Nancy E.
: APPLICANT: Harwerth, Ina-Maria
: APPLICANT: Groner, Bernd
: APPLICANT: Hardman, No. 5571894man
: APPLICANT: Zwickl, Markus
: TITLE OF INVENTION: Recombinant Antibodies Specific for a
: NUMBER OF SEQUENCES: 16
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Ciba-Geigy Corporation
: STREET: 7 Skyline Drive
: CITY: Hawthorne
: STATE: New York
: COUNTRY: USA
```



ZIP: 10532  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/235, 838  
FILING DATE: TBA  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/828, 832  
FILING DATE: 31-JAN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 91-810079.3  
FILING DATE: 05-FEB-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Elmer, James Scott  
REGISTRATION NUMBER: 36,129  
REFERENCE/DOCKET NUMBER: 4-18518/A/CIP/CONT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (919)541-8614  
TELEFAX: (919)541-8689  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2233 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Mouse and E. coli  
INDIVIDUAL ISOLATE: E. coli  
IMMEDIATE SOURCE:  
CLONE: pmw616  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 23..2155 /note= "89-445 FRP5 heavy chain  
OTHER INFORMATION: var domain; 446-490 is a linker sequence; 491-814  
OTHER INFORMATION: FRP5 light chain var domain; 815-2155 coding region  
FEATURE:  
NAME/KEY: 5'UTR  
LOCATION: 1..22  
OTHER INFORMATION: /function= "ompa 5'non-coding  
OTHER INFORMATION: region"  
FEATURE:  
NAME/KEY: sig\_peptide  
LOCATION: 23..85  
OTHER INFORMATION: /note= "ompa signal peptide"  
FEATURE:  
NAME/KEY: 3'UTR  
LOCATION: 2156..2233  
OTHER INFORMATION: /function= "phoa 3' non-coding  
OTHER INFORMATION: region"  
FEATURE:  
NAME/KEY: mat\_peptide  
LOCATION: 86..2155  
US-08-235-838-6  
Alignment Scores:  
Pred. No.: 181 Length: 2233  
Score: 43.50 Matches: 13  
Percent Similarity: 46.34% Conservative: 6  
Best Local Similarity: 31.71% Mismatches: 9  
Query Match: 33.72% Indels: 13  
DB: 1 Gaps: 1  
US-09-847-539a-6\_COPY\_59\_86 (1-28) x US-08-235-838-6 (1-2233)

Qy 1 SeraspalaLeuglAlaLeuAlaSpolnThraspala----- 13  
Db 1685 AATGACAGTGTACCAACCTGCGCCAGATGACGACAAAGCATTTGATGTAGTAA 1744  
Qy 14 -----LeuglInsergluAlaAlaValValLysAlaAspAsnAla 27  
Db 1745 AATGAGAAAGGCTTTTCCTGCAAGTTGAAGTGCATCATTAACAGATCATGCT 1804  
Qy 28 Ala 28  
Db 1805 GCG 1807  
RESULT 24  
US-08-465-473B-6  
Sequence 6, Application US/08465473B  
Patent No. 5939531  
GENERAL INFORMATION:  
APPLICANT: Wels, Manfred S.  
APPLICANT: Hynes, Nancy E.  
APPLICANT: Harwerth, Ina-Maria  
APPLICANT: Groner, Bernd  
APPLICANT: Hardman, No. 5939531man  
APPLICANT: Zwickl, Markus  
TITLE OF INVENTION: Recombinant Antibodies Specific for a  
TITLE OF INVENTION: Growth Factor Receptor  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: NOVARTIS Corporation  
STREET: 564 Morris Avenue  
CITY: Summit  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07901-6940  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/465, 473B  
FILING DATE: 5 June 1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/828, 832  
FILING DATE: 31-JAN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 91-810079.3  
FILING DATE: 05-FEB-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Pfeiffer, Henna J.  
REGISTRATION NUMBER: 22,640  
REFERENCE/DOCKET NUMBER: 4-18518/A/CIP/CONT2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (908)522 6940  
TELEFAX: (908)522 6955  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2233 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Mouse and E. coli  
INDIVIDUAL ISOLATE: E. coli  
IMMEDIATE SOURCE:  
CLONE: pmw616  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 23..2155  
OTHER INFORMATION: /note= "89-445 FRP5 heavy chain

OTHER INFORMATION: var domain; 446-490 15 aa linker sequence; 491-814  
OTHER INFORMATION: FRP5 light chain var domain; 815-2155 coding region  
OTHER INFORMATION: of phoA  
FEATURE:  
NAME/KEY: 5'UTR  
LOCATION: 1..22  
OTHER INFORMATION: /function="ompA 5' non-coding region"  
FEATURE:  
NAME/KEY: s19-peptide  
LOCATION: 23..85  
OTHER INFORMATION: /note="ompA signal peptide"  
FEATURE:  
NAME/KEY: 3'UTR  
LOCATION: 2156..2233  
OTHER INFORMATION: /function="phoA 3' non-coding region"  
OTHER INFORMATION: region"  
FEATURE:  
NAME/KEY: mat-peptide  
LOCATION: 86..2155  
US-08-465-4738-6

Alignment Scores:  
Pred. No.: 181 Length: 2233  
Score: 43.50 Matches: 13  
Percent Similarity: 46.34% Conservative: 9  
Best Local Similarity: 31.71% Mismatches: 6  
Query Match: 33.72% Indels: 13  
DB: 2 Gaps: 1

US-09-847-539a-6\_COPY\_59\_86 (1-28) x US-08-465-4738-6 (1-2233)  
QY 1 SeraspAlaLeuGluaLalaAspGlnThrAspAla----- 13  
Db 1685 AATGACAGGTGTACCAACCTGGCGGCACATGACGACCAACCAATTGATTTGTCAGTAA 1744  
QY 14 -----LeuGlnSerGluGluAlaValIaLysAlaAspAsna 27  
Db 1745 AATGAGAAAGCTTTTCTTCGCAAGTGAAGTCGTCATCGATTAACAGATCATGCT 1804  
QY 28 Ala 28  
Db 1805 GCG 1807

RESULT 25  
US-08-928-2138-7  
Sequence 7, Application US/089282138  
Patent No. 6238905  
GENERAL INFORMATION:  
APPLICANT: McHenry, Charles S.  
Seville, Mark  
Gull, Millard G.  
TITLE OF INVENTION: NOVEL THERMOPHILIC POLYMERASE III  
NUMBER OF SEQUENCES: 195  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MEDLEN & CARROLL, LLP  
STREET: 220 Montgomery Street, Suite 2200  
CITY: San Francisco  
STATE: CA  
COUNTRY: USA  
ZIP: 94104

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/928,2138  
FILING DATE: 12-Sep-1997  
CLASSIFICATION: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: MacKnight, Kamrin T.

REGISTRATION NUMBER: 38,230  
REFERENCE/DOCKET NUMBER: ENZYCO-02550  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-705-8410  
TELEFAX: 415-397-8338  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2681 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: other nucleic acid  
DESCRIPTION: /desc = "DNA"  
SEQUENCE DESCRIPTION: SEQ ID NO: 7:  
US-09-847-539a-6\_COPY\_59\_86 (1-28) x US-08-928-2138-7 (1-2681)

Alignment Scores:  
Pred. No.: 228 Length: 2681  
Score: 43.50 Matches: 12  
Percent Similarity: 73.08% Conservative: 7  
Best Local Similarity: 46.15% Mismatches: 6  
Query Match: 33.72% Indels: 1  
DB: 4 Gaps: 1

US-09-847-539a-6\_COPY\_59\_86 (1-28) x US-08-928-2138-7 (1-2681)  
QY 2 AspAlaLeuGluaLalaAspGlnThrAspAlaLeuGlnSerGluGluAlaVal 21  
Db 1686 GAGGCCATGAGAGGCGCTCGCGCCGCGCTCGACGCTTA---AGCCTGAGAGTGGCCCTC 1742  
QY 22 ValIysAlaAspAsna 27  
Db 1743 CTGAGCGCGGAGAGGCC 1760

RESULT 26  
US-08-673-312-9  
Sequence 9, Application US/08673312  
Patent No. 5891699  
GENERAL INFORMATION:  
APPLICANT: BOULAIN, JEAN-CLAUDE  
DUCANCEL, FREDERIC  
APPLICANT: MENEZ, ANDRE  
TITLE OF INVENTION: MODIFIED BACTERIAL ALKALINE PHOSPHATASES  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MATER & NEUSTADT,  
P.C.  
STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400  
CITY: ARLINGTON  
STATE: VA  
COUNTRY: USA  
ZIP: 22202

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/673,312  
FILING DATE: 28-JUN-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 95-07833  
FILING DATE: 29-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: OBLON, NORMAN F.  
REGISTRATION NUMBER: 24,618  
REFERENCE/DOCKET NUMBER: 846-380-0  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-413-3000  
TELEFAX: 703-413-2220

```
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 6162 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: double
;   TOPOLOGY: linear
;   MOLECULE TYPE: DNA (genomic)
US-08-673-12-9

Alignment Scores:
Pred. No.:      647      Length:      6162
Score:          43.50    Matches:      13
Percent Similarity: 46.34% Conservative: 6
Best Local Similarity: 31.71% Mismatches: 9
Query Match:     33.72% Indels:      13
DB:              2      Gaps:        1

US-09-847-539a-6_COPY_59_86 (1-28) x US-08-673-312-9 (1-6162)

Oy      1 SeraspAlaLeuGluaLaLeuAlaAspGlnThraspAla----- 13
Db      1252 AATGACAGTGTACCAACCTGGCCAGATGACCGACAAGCCATTGATTGTTGAGTAA 1311
Oy      14 -----LeuGlnSerGlnGluAlaAlaAlaValValLySaLaAspAsnAla 27
Db      1312 AATGAGAAAGCGTTTTCCTGCGCAAGTGAAGGTCGTCATCGATTAACAAGATCATGCT 1371
Oy      28 Ala 28
Db      1372 GCG 1374

RESULT 27
US-08-934-846-3
; Sequence 3, Application US/08934846
; Patent No. 5882898
; GENERAL INFORMATION:
;   APPLICANT: Pearson, Stewart C.
;   APPLICANT: Greenwood, Rebecca C.
;   TITLE OF INVENTION: NOVEL folC
;   NUMBER OF SEQUENCES: 4
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Dechert Price & Rhoads
;   STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
;   CITY: Philadelphia
;   STATE: PA
;   COUNTRY: US
;   ZIP: 19103
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Diskette
;   COMPUTER: IBM compatible
;   OPERATING SYSTEM: DOS
;   SOFTWARE: FASTSEQ for windows Version 2.0
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/934, 846
;   FILING DATE:
;   CLASSIFICATION: 536
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER:
;   FILING DATE:
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Dickinson, Todd O
;   REGISTRATION NUMBER: 28,354
;   REFERENCE/DOCKET NUMBER: GM10088
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 215-994-2252
;   TELEFAX: 215-994-2222
;   TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 1221 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: double
;   TOPOLOGY: linear
```

```
US-08-934-846-3

Alignment Scores:
Pred. No.:      104      Length:      1221
Score:          43.00    Matches:      9
Percent Similarity: 51.85% Conservative: 5
Best Local Similarity: 33.33% Mismatches: 13
Query Match:     33.33% Indels:      0
DB:              2      Gaps:        0

US-09-847-539a-6_COPY_59_86 (1-28) x US-08-934-846-3 (1-1221)

Oy      2 AspAlaLeuGluaLaLeuAlaAspGlnThraspAlaLeuGlnSerGlnGluAlaAlaVal 21
Db      499 GACCATCAAGAAACCTTGGGTGATGCTGAAACATTCACAGCAGAAAGCTGATTT 558
Oy      22 ValLySaLaAspAsnAlaAla 28
Db      559 TTCAAAGCGTGTAAAGAGCA 579

RESULT 28
US-09-238-557-3
; Sequence 3, Application US/09238557
; Patent No. 6165472
; GENERAL INFORMATION:
;   APPLICANT: Pearson, Stewart C.
;   APPLICANT: Greenwood, Rebecca C.
;   TITLE OF INVENTION: NOVEL folC
;   NUMBER OF SEQUENCES: 4
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Dechert Price & Rhoads
;   STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
;   CITY: Philadelphia
;   STATE: PA
;   COUNTRY: US
;   ZIP: 19103
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Diskette
;   COMPUTER: IBM compatible
;   OPERATING SYSTEM: DOS
;   SOFTWARE: FASTSEQ for windows Version 2.0
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/238,557
;   FILING DATE:
;   CLASSIFICATION:
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 08/934, 846
;   FILING DATE:
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Dickinson, Todd O
;   REGISTRATION NUMBER: 28,354
;   REFERENCE/DOCKET NUMBER: GM10088
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 215-994-2252
;   TELEFAX: 215-994-2222
;   TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 1221 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: double
;   TOPOLOGY: linear
US-09-238-557-3

Alignment Scores:
Pred. No.:      104      Length:      1221
Score:          43.00    Matches:      9
Percent Similarity: 51.85% Conservative: 5
Best Local Similarity: 33.33% Mismatches: 13
Query Match:     33.33% Indels:      0
DB:              4      Gaps:        0

US-09-847-539a-6_COPY_59_86 (1-28) x US-09-238-557-3 (1-1221)
```





```

;
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 1274 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;
; FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 179..751
;   SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-08-461-511A-1

Alignment Scores:
Pred. No.: 134 Length: 1274
Score: 42.50 Matches: 14
Percent Similarity: 65.52% Conservative: 5
Best Local Similarity: 48.28% Mismatches: 8
Query Match: 32.95% Indels: 2
DB: 1170 TGTACCCACGACATCACAGGGCC 1146 Gaps: 1

RESULT 34
PCT-US94-07089-1/c
; Sequence 1, Application PC/TUS9407089
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Vertebrate Apoptosis Gene:
; NUMBER OF SEQUENCES: 9
; ADDRESS/SEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: TX
; COUNTRY: United States of America
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS, ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/07089
; FILING DATE: CONCURRENTLY FILED
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/081,448
; FILING DATE: 22 JUNE 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: PARKER, David L.
; REGISTRATION NUMBER: 32,165
; REFERENCE/DOCKET NUMBER: ARCD090
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 512-320-7200
; TELEFAX: 713-789-2679
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 1274 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: DNA (genomic)
;   FEATURE:
;     NAME/KEY: CDS
;     LOCATION: 179..751
;     PCT-US94-07089-1
```

```

Alignment Scores:
Pred. No.: 134 Length: 1274
Score: 42.50 Matches: 14
Percent Similarity: 65.52% Conservative: 5
Best Local Similarity: 48.28% Mismatches: 8
Query Match: 32.95% Indels: 2
DB: 1170 TGTACCCACGACATCACAGGGCC 1146 Gaps: 1

US-09-847-539A-6_COPY_59_86 (1-28) x PCT-US94-07089-1 (1-1274)

QY 2 Aspa1aLeuG1a1aLeuA1a-AspGlnThrAspa1a---LeuGlnSerG1uG1a1a1 20
DB 1230 GATTCACCTCCAGCGCATGCGACGACATCTCCATCTCCAGAAATGAAGTAGC 1171

QY 20 aValValysAlaAspaSnAla1a 28
DB 1170 TGTACCCACGACATCACAGGGCC 1146

RESULT 35
US-08-081-448-1/c
; Sequence 1, Application US/08081448
; Patent No. 5646008
; GENERAL INFORMATION:
; APPLICANT: Thompson, Craig B.
; APPLICANT: Boise, Lawrence H.
; TITLE OF INVENTION: Vertebrate Apoptosis Gene:
; TITLE OF INVENTION: Compositions and Methods
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESS/SEE: Arnold, White & Durkee
; STREET: 321 No. 5646008th Clark Street, Suite 800
; CITY: Chicago
; STATE: IL
; COUNTRY: USA
; ZIP: 60610
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/081,448
; FILING DATE: 19930622
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5646008thrup, Thomas E.
; REGISTRATION NUMBER: 33,268
; REFERENCE/DOCKET NUMBER: ARCD090
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-744-0090
; TELEFAX: 312-755-4489
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 1303 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: unknown
;   MOLECULE TYPE: DNA (genomic)
;   FEATURE:
;     NAME/KEY: CDS
;     LOCATION: 179..751
;     US-08-081-448-1

Alignment Scores:
Pred. No.: 138 Length: 1303
Score: 42.50 Matches: 14
Percent Similarity: 65.52% Conservative: 5
Best Local Similarity: 48.28% Mismatches: 8
Query Match: 32.95% Indels: 2
DB: 1170 TGTACCCACGACATCACAGGGCC 1146 Gaps: 1

US-09-847-539A-6_COPY_59_86 (1-28) x US-08-081-448-1 (1-1303)
```







REGISTRATION NUMBER: 32,323  
REFERENCE/DOCKET NUMBER: CM356M  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 513-627-2885  
TELEFAX: 513-627-0318  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1060 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 10..924  
US-08-081-328-1

Alignment Scores:  
Pred. No.: 130 Length: 1060  
Score: 42.00 Matches: 15  
Percent Similarity: 59.26% Conservative: 1  
Best local Similarity: 55.56% Mismatches: 9  
Query Match: 32.56% Indels: 2  
DB: 1 Gaps: 1

US-09-847-539a-6\_COPY\_59\_86 (1-28) x US-08-081-328-1 (1-1060)

QY 2 AspaLaLeuGluAlaLeuAlaAspGlnThrAspaLaLeuGlnSerGluAlaAlaVal 21  
||||| ||||||| ||||| ||| ||| |||:|  
Db 522 GATGCCGCCGCTAGCGCTGCGCGGCGAGACCGCAACTGGGAGT-----GCATCCGTC 469  
QY 22 ValLySaLaAspaSnAlaAla 28  
||||| |||||  
Db 468 GAAGATGCCGACGCCGCCGCC 448

Search completed: October 13, 2002, 06:32:14  
Job time : 1240.98 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2002, 04:38:53 : Search time 7.63636 Seconds  
(without alignments)  
352.328 Million cell updates/sec

Title: US-09-847-539a-6\_COPY\_59\_86  
Perfect score: 129  
Sequence: 1 SDALEALADQTDALQSEEAAYVKADNAA 28

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	57.5	44.6	309	1 YAZON7	S-antigen precursor
2	57	44.2	412	2 G64685	hypothetical prote
3	57	44.2	412	2 B71831	hypothetical prote
4	51.5	39.9	213	2 JC1152	hypothetical prote
5	51	39.5	203	2 AH1080	hypothetical prote
6	51	39.5	203	2 A11437	hypothetical prote
7	51	39.5	276	2 T29125	ketocacyl reductase
8	51	39.5	677	2 S73798	MG260 homolog H91
9	51	39.5	4589	2 T14914	dynein beta heavy
10	50	38.8	617	2 G64972	yegA protein precu
11	50	38.8	617	2 G90987	suppressor of ompF
12	50	38.8	617	2 B85833	suppressor of ompF
13	50	38.8	1968	1 S05697	myosin heavy chain
14	49	38.0	142	2 AC0928	probable regulator
15	49	38.0	445	2 T06051	hypothetical prote
16	49	38.0	484	2 A84305	hypothetical prote
17	47.5	36.8	421	2 D95975	hypothetical outer
18	47	36.4	139	1 B70955	aspartate 1-decarb
19	47	36.4	192	2 T09236	nifz protein - Fra
20	47	36.4	192	2 JC4207	nifrogenase NifZ C
21	47	36.4	254	2 A71519	hypothetical prote
22	47	36.4	254	2 A81677	conserved hypotet
23	47	36.4	307	2 AB2807	conserved hypotet
24	47	36.4	333	2 A97586	hypothetical prote
25	47	36.4	1013	2 G87236	conserved integral
26	47	36.4	2429	1 S8784	spectrin alpha cha
27	46.5	36.0	536	1 D84325	Htr17 transducer
28	46.5	36.0	784	2 T51759	glutamate--ammon
29	46	35.7	106	2 A81203	conserved hypotet

30	46	35.7	123	2 E90066	aspartate 1-decarb
31	46	35.7	345	2 A87375	hypothetical prote
32	46	35.7	353	2 A72410	chorismate mutase/
33	46	35.7	425	2 T25873	hypothetical prote
34	46	35.7	526	2 C83488	probable semialden
35	46	35.7	539	2 G83720	nickel transport s
36	46	35.7	618	2 AE0770	probable outer mem
37	46	35.7	732	2 F84394	helicase [imported
38	46	35.7	764	2 A84328	Htr2 transducer [1
39	46	35.7	765	1 T44846	transducer protein
40	46	35.7	787	2 S73873	probable lipoprote
41	45.5	35.3	480	2 F81382	pyruvate kinase (E
42	45.5	35.3	557	2 AH3604	transport ATP-bind
43	45.5	35.3	572	2 T00707	transcription init
44	45.5	35.3	2186	2 H89960	hypothetical prote
45	45	34.9	84	2 E84345	hypothetical prote

## ALIGNMENTS

## RESULT 1

YAZON7  
S-antigen precursor - malaria parasite (Plasmodium falciparum) (strain NF7/Ghana)  
C:Species: Plasmodium falciparum  
C:Date: 30-Sep-1987 #sequence\_revision 30-Sep-1987 #text\_change 09-Jun-2000  
C:Accession: B22011  
R:Coman, A.F.; Saint, R.B.; Coppel, R.L.; Brown, G.V.; Anders, R.F.; Kemp, D.J.  
Cell 40, 775-783, 1985  
A:Title: Conserved sequences flank variable tandem repeats in two S-antigen genes of P.  
A:Reference number: A90863; MID:85176931  
A:Accession: B22011  
A:Molecule type: DNA  
A:Residues: 1-309 <COM>  
A:Cross-references: GB:M10130; MID:g160670; PIDN:AAA29758.1; PID:g160671  
A:Experimental source: clone NF7.S  
A:Note: the intact NF7 S-antigen contains about 35 more of the 8-residue repeats  
C:Comment: The S-antigen is secreted by the parasite.  
C:Superfamily: plasmodium S-antigen  
C:Keywords: malaria; surface antigen; tandem repeat  
F:1-23/Domain: signal sequence #status predicted <SIG>  
F:24-309/Product: S-antigen #status predicted <MAT>  
F:97-256/Region: 8-residue repeats (S-D-E-A-E-A-L/R-K)  
F:257-271,272-286/Region: 15-residue repeats

Query Match 44.6%; Score 57.5; DB 1; Length 309;  
Best local Similarity 50.0%; Pred. No. 2.1;  
Matches 14; Conservative 6; Mismatches 7; Indels 1; Gaps 1;

QY 1 SDALEALADQTDALQSEEAAYVKADNAA 27  
DB 97 SDEAELKSDAEALKSDAEARKSDEA 124

## RESULT 2

G64685  
hypothetical protein HP1327 - Helicobacter pylori (strain 26695)  
C:Species: Helicobacter pylori  
C:Date: 09-Aug-1997 #sequence\_revision 09-Aug-1997 #text\_change 08-Oct-1999  
C:Accession: G64685  
R:Tombl, J.F.; White, O.; Keriavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.  
Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Kralak, H.G.; Glodek, A.; McKee  
son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey,  
Nature 388, 539-547, 1997  
A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser,  
A:Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.  
A:Reference number: A64520; MID:97394467  
A:Accession: G64685  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-412 <COM>  
A:Cross-references: GB:AE000634; GB:AE000511; MID:g2314489; PIDN:AAD08377.1; PID:g231

```
Query Match 44.2%; Score 57; DB 2; Length 412;
Best Local Similarity 52.2%; Pred. No. 3.3;
Matches 12; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

OY 4 LEALADOTDALQSEEAAYVKADN 26
   ::||:| ||| ||| | | | |
Db 49 IQALQEQIDALDSQEKVSKMDN 71

RESULT 3
hypotheical protein jhp1247 - Helicobacter pylori (strain J99)
C:Species: Helicobacter pylori
A:Variety: Strain J99
C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 08-Oct-1999
C:Accession: B71831
R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Dolg, P.C.; Smith, D.R.;
Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;
Nature 397, 176-180, 1999
A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric path
A:Reference number: A71800; MUID:99120557
A:Accession: B71831
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-412 <ARN>
A:Cross-references: GB:AE001548; GB:AE001439; MID:g4155845; PIDN:AAD06820.1; PID:g415584
A:Experimental source: strain J99
C:Genetics:
A:Gene: jhp1247

Query Match 44.2%; Score 57; DB 2; Length 412;
Best Local Similarity 52.2%; Pred. No. 3.3;
Matches 12; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

OY 4 LEALADOTDALQSEEAAYVKADN 26
   ::||:| ||| ||| | | | |
Db 49 IQALQEQIDALDSQEKVSKMDN 71

RESULT 4
JC1152
hypotheical protein, 23.6K - Agrobacterium tumefaciens insertion sequence IS1131
C:Species: Agrobacterium tumefaciens
C:Date: 05-Mar-1993 #sequence_revision 05-Mar-1993 #text_change 17-Jul-1998
C:Accession: JC1152
R:Mabiko, H.
Gene 114, 229-233, 1992
A:Title: Sequence analysis of an insertion element, IS1131, isolated from the nopaline-t
A:Reference number: JC1150; MUID:92290280
A:Accession: JC1152
A:Molecule type: DNA
A:Residues: 1-213 <WAB>
A:Cross-references: GB:M82888
A:Experimental source: strain P022; plasmid T1; insertion sequence IS1131
C:Genetics:
A:Mobile element: insertion sequence IS1131

Query Match 39.9%; Score 51.5; DB 2; Length 213;
Best Local Similarity 52.0%; Pred. No. 9.3;
Matches 13; Conservative 5; Mismatches 6; Indels 1; Gaps 1;

OY 3 ALEAL-ADOTDALQSEEAAYVKADN 26
   || || | | | | | | | |
Db 14 ALRALVAEQAKLESQEAIVIKRDS 38

RESULT 5
AH1080
hypotheical protein lmo0047 [imported] - Listeria monocytogenes (strain EGD-e)
C:Species: Listeria monocytogenes
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
C:Accession: AH1080
R:Glasger, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker
```

```
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihl,
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Kretz, J.; Kuhn, M.; Kunst, F.; Kurapat, G.; Madueno, E.; Maitournam, A.;
Ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehla
A:Title: Comparative genomics of Listeria species
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AH1080
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-203 <GLA>
A:Cross-references: GB:NC_003210; PIDN:GAC98262.1; PID:g16409406; GSPDB:GN00177
A:Experimental source: strain EGD-e
C:Genetics:
A:Gene: lmo0047

Query Match 39.5%; Score 51; DB 2; Length 203;
Best Local Similarity 38.5%; Pred. No. 10;
Matches 10; Conservative 7; Mismatches 9; Indels 0; Gaps 0;

OY 1 SDALEADOTDALQSEEAAYVKADN 26
   :: ||| ||| | | | | |
Db 109 AETKEQLSDETDRDREDAGVEXEN 134

RESULT 6
AH1437
hypotheical protein lln0040 [imported] - Listeria innocua (strain C1p11262)
C:Species: Listeria innocua
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
C:Accession: AH1437
R:Glasger, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloec
D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihl,
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Kretz, J.; Kuhn, M.; Kunst, F.; Kurapat, G.; Madueno, E.; Maitournam, A.;
Ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehla
A:Title: Comparative genomics of Listeria species
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AH1437
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-203 <GLA>
A:Cross-references: GB:AL592022; PIDN:CAC95273.1; PID:g16412461; GSPDB:GN00178
A:Experimental source: strain C1p11262
C:Genetics:
A:Gene: lln0040

Query Match 39.5%; Score 51; DB 2; Length 203;
Best Local Similarity 38.5%; Pred. No. 10;
Matches 10; Conservative 6; Mismatches 10; Indels 0; Gaps 0;

OY 1 SDALEADOTDALQSEEAAYVKADN 26
   :: ||| ||| | | | | |
Db 109 AETKELEADETDKLDEKGVEXEN 134

RESULT 7
T29125
ketocacyl reductase homolog - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 20-Jun-2000
C:Accession: T29125
R:Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, August 1998
A:Reference number: Z17215
A:Accession: T29125
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-276 <PAR>
A:Cross-references: EMBL:AL031350; PIDN:CAA20507.1
C:Genetics:
A:Note: SC1P2.16c
```







RESULT 20  
JC4207  
N: nitrogenase Nifz chain - Frankia sp.  
N: Alternate names: Nifz protein  
C: Species: Frankia sp.  
C: Date: 10-Sep-1995 #sequence\_revision 27-Oct-1995 #text\_change 22-Oct-1999  
C: Accession: JC4207  
R: Harriott, O.T.; Hosted, T.J.; Benson, D.R.  
Gene 161, 63-67, 1995  
A: Title: Sequences of nifx, nifw, nifz, nifB and two ORF in the Frankia nitrogen fixative  
A: Reference numbers: JC4203; MUID:95369734  
A: Accession: JC4207  
A: Molecule type: DNA  
A: Residues: 1-192 <HAR>  
A: Cross-references: GB:L29299; NID:9497430; PIDN:AAC82974.1; PID:9497435  
A: Comment: This protein has an Ala and Pro rich region at its carboxyl-terminal region.  
C: Genetics:  
A: Gene: nifz  
C: Keywords: nitrogen fixation

Query Match 36.4%; Score 47; DB 2; Length 192;  
Best Local Similarity 52.0%; Pred. No. 35;  
Matches 13; Conservative 1; Mismatches 11; Indels 0; Gaps 0;

OY 4 LEALADOTDALQSEAAVVRADNA 28  
Db 74 LEAAEDADATQAPPAATPTAAAAA 98

RESULT 21  
A71519  
hypothetical protein CT398 - Chlamydia trachomatis (serotype D, strain UW3/Cx)  
C: Species: Chlamydia trachomatis  
C: Date: 13-Sep-1998 #sequence\_revision 13-Sep-1998 #text\_change 08-Oct-1999  
C: Accession: A71519  
R: Stephens, R.S.; Kaiman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell,  
Science 282, 754-759, 1998  
A: Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trac  
A: Reference number: A71570; MUID:9900809  
A: Accession: A71519  
A: Status: preliminary  
A: Molecule type: DNA  
A: Residues: 1-254 <ARN>  
A: Cross-references: GB:AE001333; GB:AE001273; NID:93328823; PIDN:AAC67995.1; PID:9332882  
A: Experimental source: serotype D, strain UW-3/Cx  
C: Genetics:  
A: Gene: CT398

Query Match 36.4%; Score 47; DB 2; Length 254;  
Best Local Similarity 36.4%; Pred. No. 46;  
Matches 8; Conservative 9; Mismatches 5; Indels 0; Gaps 0;

OY 4 LEALADOTDALQSEAAVVRAD 25  
Db 67 IQEISDQINKLENOQAARVKKMD 88

RESULT 22  
A81677  
conserved hypothetical protein TC0677 [Imported] - Chlamydia muridarum (strain N19g)  
C: Species: Chlamydia muridarum, Chlamydia trachomatis Mopn  
C: Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 11-May-2000  
C: Accession: A81677  
R: Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,  
C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg,  
Nucleic Acids Res. 28, 1397-1406, 2000  
A: Title: Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39.  
A: Reference number: A81500; MUID:20150255  
A: Accession: A81677  
A: Status: preliminary  
A: Molecule type: DNA  
A: Residues: 1-254 <TET>

A: Cross-references: GB:AE002336; GB:AE002160; NID:97190707; PIDN:AAE39497.1; PID:9719  
A: Experimental source: strain N19g (Mopn)  
C: Genetics:  
A: Gene: TC0677

Query Match 36.4%; Score 47; DB 2; Length 254;  
Best Local Similarity 36.4%; Pred. No. 46;  
Matches 8; Conservative 9; Mismatches 5; Indels 0; Gaps 0;

OY 4 LEALADOTDALQSEAAVVRAD 25  
Db 67 IQEISDQINKLENOQAARVKKMD 88

RESULT 23  
AB2807  
conserved hypothetical protein Atu1876 [Imported] - Agrobacterium tumefaciens (strain  
C: Species: Agrobacterium tumefaciens  
C: Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 11-Jan-2002  
C: Accession: AB2807  
R: Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo  
erage, G.; Giller, W.; Grant, C.; Guenther, D.; Kuttyavin, T.; Levy, R.; Li, M.; McL  
; Karp, P.; Romero, P.; Zhang, S.  
Science 294, 2317-2323, 2001  
A: Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kam  
ster, E.W.  
A: Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
A: Reference number: AB2577; PMID:11743193  
A: Accession: AB2807  
A: Status: preliminary  
A: Molecule type: DNA  
A: Residues: 1-307 <KUR>  
A: Cross-references: GB:AE008688; PIDN:AAL42872.1; PID:917740324; GSPDB:GNO0186  
A: Experimental source: strain C58 (Dupont)  
C: Genetics:  
A: Gene: Atu1876  
A: Map position: circular chromosome

Query Match 36.4%; Score 47; DB 2; Length 307;  
Best Local Similarity 47.8%; Pred. No. 57;  
Matches 11; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

OY 2 DALEALADOTDALQSEAAVYKA 24  
Db 79 DQLEALADOTDALQSEAAVYKA 101

RESULT 24  
A97586  
hypothetical protein AGR\_C\_3443 [Imported] - Agrobacterium tumefaciens (strain C58, C  
C: Species: Agrobacterium tumefaciens  
C: Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 11-Jan-2002  
C: Accession: A97586  
R: Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Gold  
A.; Liu, F.; Mollam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Leppas, C.; Markelz,  
Science 294, 2323-2328, 2001  
A: Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium  
A: Reference number: A97359; PMID:11743194  
A: Accession: A97586  
A: Status: preliminary  
A: Molecule type: DNA  
A: Residues: 1-333 <KUR>  
A: Cross-references: GB:AE007869; PIDN:AAK87642.1; PID:915156994; GSPDB:GNO0169  
C: Genetics:  
A: Gene: AGR\_C\_3443  
A: Map position: circular chromosome

Query Match 36.4%; Score 47; DB 2; Length 333;  
Best Local Similarity 47.8%; Pred. No. 62;  
Matches 11; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

OY 2 DALEALADOTDALQSEAAVYKA 24  
Db 79 DQLEALADOTDALQSEAAVYKA 101



```
Db 105 DQLEALQTTDAMKVRADIAFA 127
RESULT 25
conserved integral membrane protein M2617 [imported] - Mycobacterium leprae
C:Species: Mycobacterium leprae
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C:Accession: G87236
R: Cole, S.T.; Eislmeier, K.; Parkhill, J.; James, K.D.; Thomson, M.R.; Wheeler, P.R.; Ho
R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Fellwell, T.; Fraser, A.; Hamlin, N.; Holroyd,
eam, M.A.; Rutherford, K.M.
Nature 409, 1007-1011, 2001
A: Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Se
A: Title: Massive gene decay in the leprosy bacillus.
A: Reference number: A86909; MUID:21128732; PMID:11234002
A: Accession: G87236
A: Status: preliminary
A: Molecule type: DNA
A: Residues: 1-1013 <STO>
A: Cross-references: GB:AL450380; NID:g13093843; PIDN:CAC32149.1; GSPDB:GN00147
C: Genetics:
A: Gene: M2617

Query Match 36.4%; Score 47; DB 2; Length 1013;
Best Local Similarity 62.5%; Pred. No. 2e+02;
Matches 10; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Oy 7 LADQTDALQSEAAVY 22
Db 931 LADWDALIEPTAAVY 946

RESULT 26
SJHUA
Spectrin alpha chain - human
C:Species: Homo sapiens (man)
C:Date: 17-May-1985 #sequence_revision 30-Jun-1992 #text_change 16-Jul-1999
C:Accession: A35716; A45755; A23533; A91528; A42872; A02965; S13138; A05282; A23965; A27
R: Sahr, K.E.; Laurila, P.; Kotila, L.; Scarpa, A.L.; Coupal, E.; Leco, T.L.; Linnenbach,
J. Biol. Chem. 265, 4434-4443, 1990
A: Title: The complete cDNA and polypeptide sequences of human erythroid alpha-spectrin.
A: Reference number: A35716; MUID:90170949
A: Accession: A35716
A: Molecule type: mRNA
A: Residues: 1-2429 <SAH>
A: Cross-references: GB:J05244
R: Sahr, K.E.; Tobe, T.; Scarpa, A.; Laughinghouse, K.; Marchesi, S.L.; Agre, P.; Linnet
J. Clin. Invest. 84, 1243-1252, 1989
A: Title: Sequence and exon-intron organization of the DNA encoding the alpha domain of
A: Reference number: A45755; MUID:90003318
A: Accession: A45755
A: Molecule type: DNA
A: Residues: 1-394; G', 396-533 <SAZ>
A: Cross-references: GB:M29983
A: Note: the authors translated the codon GGT for residue 395 as Ala
R: Linnenbach, A.J.; Speicher, D.W.; Marchesi, V.T.; Forget, B.G.
Proc. Natl. Acad. Sci. U.S.A. 83, 2397-2401, 1986
A: Title: Cloning of a portion of the chromosomal gene for human erythrocyte alpha-spectr
A: Reference number: A23533; MUID:86205962
A: Accession: A23533
A: Molecule type: DNA
A: Residues: 320-450 <LIN>
A: Cross-references: GB:M13233; NID:g182242; PIDN:AAA53103.1; PID:g182243
R: Curtis, P.J.; Palumbo, A.; Ming, J.; Fraser, P.; Cioe, L.; Meo, P.; Shane, S.; Rovera,
Gene 36, 357-362, 1985
A: Title: Sequence comparison of human and murine erythrocyte alpha-spectrin cDNA.
A: Reference number: A91528; MUID:86083178
A: Accession: A91528
A: Molecule type: mRNA
A: Residues: 1451-1687 <CUR>
A: Cross-references: GB:M11049; NID:g338310; PIDN:AAA60569.1; PID:g553648
R: Speicher, D.W.; Weglartz, L.; Desilva, T.M.

J. Biol. Chem. 267, 14775-14782, 1992
A: Title: Properties of human red cell spectrin heterodimer (side-to-side) assembly an
A: Reference number: A42872; MUID:92340516
A: Accession: A42872
A: Molecule type: protein
A: Residues: 7-16; 46-55; 680-689; 1047-1056; 1921-1930 <SPI>
R: Speicher, D.W.; Davis, G.; Marchesi, V.T.
J. Biol. Chem. 258, 14938-14947, 1983
A: Title: Structure of human erythrocyte spectrin. II. The sequence of the alpha-I dom
A: Reference number: A92408; MUID:8408788
A: Accession: A02965
A: Molecule type: protein
A: Residues: 7-601 <SPC>
R: Speicher, D.W.; Davis, G.; Vurchenco, P.D.; Marchesi, V.T.
J. Biol. Chem. 258, 14931-14937, 1983
A: Title: Structure of human erythrocyte spectrin. I. Isolation of the alpha-I domain
A: Reference number: S13138; MUID:84087887
A: Accession: S13138
A: Molecule type: protein
A: Residues: 7-92; 'X', 94-96, 110-151; 317-342; 345-366, 'Z', 368-370, 'X', 372-373; 387-434; 45
R: Lusitani, D.M.; Qataishat, N.; Labraque, C.C.; Yu, R.N.; Davis, J.; Kelley, M.R.; Fun
J. Biol. Chem. 269, 25955-25958, 1994
A: Title: The first human alpha-spectrin structural domain begins with serine.
A: Reference number: A38928; MUID:95014412
A: Contents: annotation; determination of structural domain
A: Comment: Spectrin associates with band 4.1 and actin to form the cytoskeletal super
S of approximately 106 residues each.
C: Genetics:
A: Gene: GDB:SPR1
A: Cross-references: GDB:119601; OMIM:182860
A: Map position: 1q21-1q21
C: Superfamily: spectrin alpha chain; calmodulin repeat homology; SH3 homology; spectr
C: Keywords: actin binding; cytoskeleton; duplication; EF hand; erythrocyte; heterodim
F: 52-156/Domain: spectrin/dystrophin repeat homology <SP1>
F: 157-262/Domain: spectrin/dystrophin repeat homology <SP2>
F: 263-368/Domain: spectrin/dystrophin repeat homology <SP3>
F: 369-474/Domain: spectrin/dystrophin repeat homology <SP4>
F: 475-580/Domain: spectrin/dystrophin repeat homology <SP5>
F: 581-665/Domain: spectrin/dystrophin repeat homology <SP6>
F: 666-791/Domain: spectrin/dystrophin repeat homology <SP7>
F: 792-897/Domain: spectrin/dystrophin repeat homology <SP8>
F: 898-993/Domain: spectrin/dystrophin repeat homology #status atypical <SP9>
F: 994-1031/Domain: SH3 homology <SH3>
F: 1081-1181/Domain: spectrin/dystrophin repeat homology <SP10>
F: 1182-1287/Domain: spectrin/dystrophin repeat homology <SP11>
F: 1288-1393/Domain: spectrin/dystrophin repeat homology <SP12>
F: 1394-1498/Domain: spectrin/dystrophin repeat homology <SP13>
F: 1499-1604/Domain: spectrin/dystrophin repeat homology <SP14>
F: 1605-1710/Domain: spectrin/dystrophin repeat homology <SP15>
F: 1711-1816/Domain: spectrin/dystrophin repeat homology <SP16>
F: 1817-1925/Domain: spectrin/dystrophin repeat homology <SP17>
F: 1926-2032/Domain: spectrin/dystrophin repeat homology <SP18>
F: 2041-2146/Domain: spectrin/dystrophin repeat homology <SP19>
F: 2155-2257/Domain: spectrin/dystrophin repeat homology <SP20>
F: 2270-2302/Domain: calmodulin repeat homology <EF2>
F: 2313-2345/Domain: calmodulin repeat homology <EF2>

Query Match 36.4%; Score 47; DB 1; Length 2429;
Best Local Similarity 40.9%; Pred. No. 5e+02;
Matches 9; Conservative 8; Mismatches 5; Indels 0; Gaps 0;

Oy 2 DALEALDQTDALQSEAAVY 23
Db 953 DSMKALRNQANACOOQAAVPE 974

RESULT 27
DB4325
Htt17 transducer [imported] - Halobacterium sp. NRC-1
C:Species: Halobacterium sp. NRC-1
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
C:Accession: DB4325
R: Ng, W.V.; Kennedy, S.P.; Mahalires, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky
```

; Leithauer, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablo  
Jung, K.H.; Alam, M.; Freltas, T.  
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000  
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Edhardt, H.; Lowe, T.M.; Li  
A:Title: Genome sequence of *Halobacterium* species NRC-1.  
A:Reference number: A84160; MUID:20504483  
A:Accession: D84325  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-536 <STO>  
A:Cross-references: GB:AE004437; NID:g10581193; PIDN:AA619968.1; GSPDB:GN00138  
C:Genetics:  
A:Gene: htr17  
C:Superfamily: Halobacterium salinarum transducer protein htr1  
Query Match 36.0%; Score 46.5; DB 2; Length 536;  
Best Local Similarity 50.0%; Pred. No. 1.2e+02;  
Matches 13; Conservative 2; Mismatches 10; Indels 1; Gaps 1;  
OY 2 DALEALADQTDALQSEEAAYVAKDNA 27  
DB 295 DAAAAVTQMTD-IQSETEAVVSEANA 319  
RESULT 28  
T51759  
[glutamate--ammonia-lyase] adenylyltransferase (EC 2.7.7.42) [Imported] - Streptomyces  
C:Species: Streptomyces coelicolor  
C>Date: 18-Aug-2000 #sequence\_revision 18-Aug-2000 #text\_change 18-Aug-2000  
C:Accession: T51759  
R:Pink, D.; Falke, D.; Wohlleben, W.; Engels, A.  
Microbiology 145, 2313-2322, 1999  
A:Title: Nitrogen metabolism in *Streptomyces coelicolor* A3(2): modification of glutamine  
A:Reference number: Z25448  
A:Accession: T51759  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-784 <PIN>  
A:Cross-references: EMBL:Y17736; PIDN:CAA76840.2  
C:Genetics:  
A:Gene: glnE  
C:Keywords: nucleotidyltransferase  
Query Match 36.0%; Score 46.5; DB 2; Length 784;  
Best Local Similarity 56.5%; Pred. No. 1.8e+02;  
Matches 13; Conservative 2; Mismatches 3; Indels 5; Gaps 1;  
OY 2 DALEALA-----DQTDALQSEEA 19  
DB 391 DALEALACGYGRTDAALQDEA 413  
RESULT 29  
A81203  
conserved hypothetical protein NMB0404 [Imported] - *Neisseria meningitidis* (strain MC58  
C:Species: *Neisseria meningitidis*  
C>Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 19-Jan-2001  
C:Accession: A81203  
R:Teitelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.  
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;  
Li, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignan, V.; Pizzi, M.  
Science 287, 1809-1815, 2000  
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ve  
A:Title: Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58.  
A:Reference number: A81000; MUID:20175755  
A:Accession: A81203  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-106 <TEI>  
A:Cross-references: GB:AE002396; GB:AE002098; NID:g7225622; PIDN:AAF40843.1; PID:g722562  
C:Genetics:  
A:Experimental source: serogroup B, strain MC58  
A:Gene: NMB0404

Query Match 35.7%; Score 46; DB 2; Length 106;  
Best Local Similarity 42.3%; Pred. No. 25;  
Matches 11; Conservative 6; Mismatches 9; Indels 0; Gaps 0;  
OY 3 ALEALADQTDALQSEEAAYVAKDNA 28  
DB 65 ALEALRELEAQNPERALEAAEA 90  
RESULT 30  
E90066  
aspartate 1-decarboxylase [Imported] - *Staphylococcus aureus* (strain N315)  
C:Species: *Staphylococcus aureus*  
C>Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 22-Oct-2001  
C:Accession: E90066  
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; O  
ma, A.; Mizutani-Ul, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K  
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiratsugu, K.  
Lancet 357, 1225-1240, 2001  
A:Title: Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*.  
A:Reference number: A89758; MUID:21311952; PMID:11418146  
A:Accession: E90066  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-123 <KUR>  
A:Cross-references: GB:BA000018; PID:g13702554; PIDN:BA843695.1; GSPDB:GN00149  
A:Experimental source: strain N315  
C:Genetics:  
A:Gene: pand  
C:Superfamily: aspartate 1-decarboxylase  
Query Match 35.7%; Score 46; DB 2; Length 123;  
Best Local Similarity 42.9%; Pred. No. 30;  
Matches 12; Conservative 5; Mismatches 7; Indels 4; Gaps 1;  
OY 1 SDALALADQTDALQSEEAAYVAKDNA 28  
DB 26 SDILEAV---DILPEKVAIVNNNGA 49  
RESULT 31  
A87375  
hypothetical protein CC1013 [Imported] - *Caulobacter crescentus*  
C:Species: *Caulobacter crescentus*  
C>Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001  
C:Accession: A87375  
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg,  
B.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwin, M.L.; Haft, D.H.; Ko  
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C  
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
A:Title: Complete genome sequence of *Caulobacter crescentus*.  
A:Reference number: A87249; MUID:21173698; PMID:11259647  
A:Accession: A87375  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-345 <STO>  
A:Cross-references: GB:AE005673; NID:g13422303; PIDN:AAK22997.1; GSPDB:GN00148  
C:Genetics:  
A:Gene: CC1013  
Query Match 35.7%; Score 46; DB 2; Length 345;  
Best Local Similarity 43.5%; Pred. No. 88;  
Matches 10; Conservative 5; Mismatches 8; Indels 0; Gaps 0;  
OY 6 ALADQTDALQSEEAAYVAKDNA 28  
DB 303 ALAEQFDRLHADKAATAKMGTA 325  
RESULT 32  
A72410  
chorismate mutase/prephenate dehydratase - *Thermotoga maritima* (strain MSB8)





GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

## SUMMARIES

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 13, 2002, 04:48:12 : Search time 39.5294 Seconds

(without alignments)  
1216.148 Million cell updates/sec

Title: US-09-847-539a-6\_COPY\_59\_86  
Perfect score: 129  
Sequence: 1 SDLEALADQTDALQSEEAAYKADNNA 28

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1736436 segs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

## Command line parameters:

-MODEL-frame+ .p2n.model -DEV-x1h  
-O/cgnt2\_1/USPRO.spool/US09847539/runtat\_10102002\_093105\_4982/app-query.fasta.1.526  
-DB-N-geneseq\_032802 -QPMT-fastlap -SUFFIX-std.rng -MINMATCH=0.1 -LOOPEXT=0  
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=40  
-MODE-LOCAL -OUTPMT-pto -NORM-ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000  
-USR-US09847539 -CGCN\_1\_1\_48 -tunat\_10102002\_093105\_4982 -NCPU=6 -ICPU=3  
-NO\_XLXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

## Database :

N\_Geneseq\_032802: \*  
1: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1980.DAT: \*  
2: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1981.DAT: \*  
3: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1982.DAT: \*  
4: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1983.DAT: \*  
5: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1984.DAT: \*  
6: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1985.DAT: \*  
7: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1986.DAT: \*  
8: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1987.DAT: \*  
9: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1988.DAT: \*  
10: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1989.DAT: \*  
11: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1990.DAT: \*  
12: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1991.DAT: \*  
13: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1992.DAT: \*  
14: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1993.DAT: \*  
15: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1994.DAT: \*  
16: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1995.DAT: \*  
17: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1996.DAT: \*  
18: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT: \*  
19: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1998.DAT: \*  
20: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1999.DAT: \*  
21: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT: \*  
22: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT: \*  
23: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT: \*  
24: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2002.DAT: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

Result				SUMMARIES			
No.	Score	Query Match	Length	DB ID	Description		
1	129	100.0	469	21	ADD00562		
2	129	100.0	504	21	ADD00564		
3	129	100.0	654	21	ADD00560		
4	129	100.0	764	21	ADD00559		
5	129	100.0	777	21	ADD00561		
6	129	100.0	853	21	ADD00563		
7	57	44.2	576	18	AA767618		
8	57	44.2	666	18	AA767891		
9	57	44.2	1239	19	AAK30461		
10	54	41.9	3349	23	ABLI3527		
11	54	41.9	6398	23	ABLI3526		
12	52	40.3	1002	21	AAK54228		
13	52	40.3	1230	21	AAK33266		
14	52	40.3	16135	23	ABLI25211		
15	52	40.3	18603	23	ABLI25210		
16	51.5	39.9	25461	23	ABLI01876		
17	51	39.5	1960	23	ABLI2683		
18	51	39.5	2618	23	ABLI27360		
19	51	39.5	4221	23	ABLI2682		
20	51	39.5	2944528	24	ABK03041		
21	50	38.8	3711	23	AAK54455		
22	48	37.2	1616	23	ABLI25849		
23	48	37.2	4229	23	ABLI25848		
24	47	36.4	4403765	22	AAI99683		
25	47	36.4	4411529	22	AAI99682		
26	46	35.7	636	21	AAK1343		
27	46	35.7	2115	22	AAK65520		
28	46	35.7	2160	22	AAK65521		
29	46	35.7	4549	18	AAK74485		
30	46	35.7	69936	21	AAK1479		
31	46	35.7	349980	21	AAK21607		
32	46	35.7	349980	22	AAK68525		
33	46	35.7	1437668	21	AAK81490		
34	45.5	35.3	234	22	ABA74678		
35	45.5	35.3	234	22	AAK23154		
36	45.5	35.3	234	22	AAK49325		
37	45.5	35.3	234	22	AAI55172		
38	45.5	35.3	544	22	ABA62184		
39	45.5	35.3	544	22	AAK10508		
40	45.5	35.3	544	22	AAK36405		
41	45.5	35.3	544	22	AAI42139		
42	45.5	35.3	1719	21	AAK43055		
43	45.5	35.3	6228	23	AAK52178		
44	45.5	35.3	6561	23	AAK55178		
45	45.5	35.3	8009	18	AAK74372		

ALIGNMENTS

RESULT 1

ADD00562

ADD00562 standard; DNA; 469 BP.

29-AUG-2000 (first entry)

Streptococcus pyogenes strain API partial GRAB protein encoding DNA.

GRAB protein: protein G related alpha2M binding protein; vaccine;

alpha2-macroglobulin: group A Streptococcus; GAS; antibiotic;

immune response; Streptococcus pyogenes infection; ds.

Streptococcus pyogenes.

Location/Qualifiers

1..468

CDS

```
FT      /tag= a
FT      /product= "GRAB protein"
FT      /partial
XX
XX      WO200026240-A2.
XX
XX      11-MAY-2000.
XX
XX      02-NOV-1999: 99WO-GB03631.
XX
XX      02-NOV-1998: 98GB-0023975.
XX
XX      (ACTI-) ACTINOVA LTD.
XX
XX      Bjorck LH, Rasmussen M;
XX
XX      WPI: 2000-365572/31.
XX      P-PSDB: AAY71044.
XX
XX      New alpha2M binding protein for generating a protective immune response
XX      to group A streptococcus and purifying the binding protein -
XX
XX      Claim 13; Page 64; 67pp: English.
XX
XX      The patent discloses a new family of proteins termed GRAB (protein G
XX      related alpha2M binding protein) from streptococcus pyogenes which have
XX      the ability to bind alpha2-macroglobulin (alpha2M) and show homology to
XX      protein G of group G streptococcus. GRAB protein and peptides derived
XX      from it are used in vaccine compositions for generating a protective
XX      immune response against group A streptococcus. Antibodies against GRAB
XX      are useful for treating streptococcus pyogenes infections. The protein
XX      is also useful for purifying alpha2M from a sample. The present sequence
XX      is a DNA encoding partial GRAB protein from S. pyogenes strain A1.
XX      The protein has alpha2M binding region and is useful in vaccine
XX      composition.
XX
XX      SQ      Sequence 469 BP; 180 A; 90 C; 99 G; 100 T; 0 other;
XX
XX      Alignment Scores:
XX      Pred. No.: 4,46e-13      Length: 469
XX      Score: 129.00      Matches: 28
XX      Percent Similarity: 100.00%      Conservative: 0
XX      Best Local Similarity: 100.00%      Mismatches: 0
XX      Query Match: 100.00%      Indels: 0
XX      DB: 21      Gaps: 0
XX
XX      US-09-847-539A-6_COPY_59_86 (1-28) x AAD00562 (1-469)
XX
XX      QY      1 SerASPAlaLeuGluAlaLeuAlaAspGlnThrASPAlaLeuGlnSerGluGluAlaAla 20
XX      |||||||
XX      DB      172 TCAGATGCGCTTAGAACATTAGCGGATCAACAGACGCTTTACAAATCAGAAAGACCTGCG 231
XX
XX      QY      21 ValValIysAlaAspAsnAlaAla 28
XX      |||||||
XX      DB      232 GTTGTTAAAGCGGATTAACGCTGCT 255
XX
XX      RESULT 2
XX      AAD00564
XX      ID      AAD00564 standard; DNA: 504 BP.
XX
XX      AC      AAD00564;
XX
XX      DT      29-AUG-2000 (first entry)
XX
XX      DE      Streptococcus pyogenes strain KTL3 partial GRAB protein encoding DNA.
XX
XX      KW      GRAB protein; protein G related alpha2M binding protein; vaccine;
XX      alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;
XX      immune response; Streptococcus pyogenes infection; ds.
XX
XX      OS      Streptococcus pyogenes.
XX
XX      FH      Key      Location/Qualifiers
XX
```

```
FT      CDS      1..504
FT      /tag= a
FT      /product= "GRAB protein"
FT      /partial
XX
XX      WO200026240-A2.
XX
XX      11-MAY-2000.
XX
XX      02-NOV-1999: 99WO-GB03631.
XX
XX      02-NOV-1998: 98GB-0023975.
XX
XX      (ACTI-) ACTINOVA LTD.
XX
XX      Bjorck LH, Rasmussen M;
XX
XX      WPI: 2000-365572/31.
XX      P-PSDB: AAY71046.
XX
XX      New alpha2M binding protein for generating a protective immune response
XX      to group A streptococcus and purifying the binding protein -
XX
XX      Claim 13; Page 65; 67pp: English.
XX
XX      The patent discloses a new family of proteins termed GRAB (protein G
XX      related alpha2M binding protein) from streptococcus pyogenes which have
XX      the ability to bind alpha2-macroglobulin (alpha2M) and show homology to
XX      protein G of group G streptococcus. GRAB protein and peptides derived
XX      from it are used in vaccine compositions for generating a protective
XX      immune response against group A streptococcus. Antibodies against GRAB
XX      are useful for treating streptococcus pyogenes infections. The protein
XX      is also useful for purifying alpha2M from a sample. The present sequence
XX      is a DNA encoding partial GRAB protein from S. pyogenes strain KTL3.
XX      The protein has alpha2M binding region and is useful in vaccine
XX      composition.
XX
XX      SQ      Sequence 504 BP; 188 A; 97 C; 108 G; 111 T; 0 other;
XX
XX      Alignment Scores:
XX      Pred. No.: 4,89e-13      Length: 504
XX      Score: 129.00      Matches: 28
XX      Percent Similarity: 100.00%      Conservative: 0
XX      Best Local Similarity: 100.00%      Mismatches: 0
XX      Query Match: 100.00%      Indels: 0
XX      DB: 21      Gaps: 0
XX
XX      US-09-847-539A-6_COPY_59_86 (1-28) x AAD00564 (1-504)
XX
XX      QY      1 SerASPAlaLeuGluAlaLeuAlaAspGlnThrASPAlaLeuGlnSerGluGluAlaAla 20
XX      |||||||
XX      DB      208 TCAGATGCGCTTAGAACATTAGCGGATCAACAGACGCTTTACAAATCAGAAAGACCTGCG 267
XX
XX      QY      21 ValValIysAlaAspAsnAlaAla 28
XX      |||||||
XX      DB      268 GTTGTTAAAGCGGATTAACGCTGCT 291
XX
XX      RESULT 3
XX      AAD00560
XX      ID      AAD00560 standard; DNA: 654 BP.
XX
XX      AC      AAD00560;
XX
XX      DT      29-AUG-2000 (first entry)
XX
XX      DE      Streptococcus pyogenes strain SF370 GRAB protein coding region.
XX
XX      KW      GRAB protein; protein G related alpha2M binding protein; vaccine;
XX      alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;
XX      immune response; Streptococcus pyogenes infection; ds.
XX
XX      OS      Streptococcus pyogenes.
XX
XX      FH      Key      Location/Qualifiers
XX
```

```

FH Key          Location/Qualifiers
FT CDS           1..654
FT              /*tag- a
FT              /product= "GRAB protein"
FT sig-peptide   1..99
FT              /*tag- b
FT mat_peptide   100..651
FT              /*tag- c
FT              /product= "Mature GRAB protein"
XX
PN WO200026240-A2.
XX
PD 11-MAY-2000.
XX
PF 02-NOV-1999; 99MO-GB03631.
XX
PR 02-NOV-1998; 98GB-0023975.
XX
PA (ACT1-) ACTINOVA LTD.
XX
PI Bjorck LH, Rasmussen M;
XX
DR WPI, 2000-365572/31.
DR P-PSDB; AAY71042.
XX
XX
PT New alpha2M binding protein for generating a protective immune response
PT to group A streptococcus and purifying the binding protein
XX
XX
PS Claim 13; Page 63; 67pp; English.
XX
CC The patent discloses a new family of proteins termed GRAB (protein G
CC related alpha2M binding protein) from Streptococcus pyogenes which have
CC the ability to bind alpha2-macroglobulin (alpha2M) and show homology to
CC protein G of group G Streptococcus. GRAB protein and peptides derived
CC from it are used in vaccine compositions for generating a protective
CC immune response against group A Streptococcus. Antibodies against GRAB
CC are useful for treating Streptococcus pyogenes infections. The protein
CC is also useful for purifying alpha2M from a sample. The present sequence
CC is a DNA encoding full-length GRAB protein from S. pyogenes strain SF370.
XX
SQ Sequence 654 BP; 234 A; 118 C; 143 G; 159 T; 0 other;
XX
Alignment Scores:
Pred. No.: 6,79e-13 Length: 654
Score: 129.00 Matches: 28
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Gaps: 0
DB: 21
XX
US-09-847-539a-6_COPY_59_86 (1-28) x AAD00560 (1-654)
QY 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluAlaAla 20
Db 274 TCAGATGCGCTTAGAAGCATTTAGCGGATCAACAGACGCTTACATCAATGAGAGAGCTGCG 333
QY 21 ValValysAlaAspAsnAlaAla 28
Db 334 GTTGTTAAAGCGATTAACGCTGCT 357
XX
RESULT 4
AAD00559
ID AAD00559 standard; DNA; 764 BP.
AC AAD00559;
XX
XX
DT 29-AUG-2000 (first entry)
XX
DE Streptococcus pyogenes strain SF370 GRAB protein encoding DNA.
XX
KW GRAB protein; protein G related alpha2M binding protein; vaccine;
KW alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;
KW immune response; Streptococcus pyogenes infection; ds.
```

```

XX
XX Streptococcus pyogenes.
OS
XX
FH Key          Location/Qualifiers
FT CDS           60..713
FT              /*tag- a
FT sig-peptide   60..158
FT              /*tag- b
FT mat_peptide   159..710
FT              /*tag- c
FT              /product= "Mature GRAB protein"
FT primer_bind   complement (101..124)
FT              /*tag- d
FT primer_bind   complement (101..127)
FT              /*tag- e
FT primer_bind   complement (160..184)
FT              /*tag- f
FT primer_bind   complement (363..594)
FT              /*tag- g
FT primer_bind   complement (565..626)
FT              /*tag- h
FT primer_bind   complement (565..626)
XX
XX WO200026240-A2.
XX
PD 11-MAY-2000.
XX
PF 02-NOV-1999; 99MO-GB03631.
XX
PR 02-NOV-1998; 98GB-0023975.
XX
PA (ACT1-) ACTINOVA LTD.
XX
PI Bjorck LH, Rasmussen M;
XX
DR WPI, 2000-365572/31.
DR P-PSDB; AAY71042.
XX
XX
PT New alpha2M binding protein for generating a protective immune response
PT to group A streptococcus and purifying the binding protein
XX
XX
PS Example 1; Fig 2B; 67pp; English.
XX
CC The patent discloses a new family of proteins termed GRAB (protein G
CC related alpha2M binding protein) from Streptococcus pyogenes which have
CC the ability to bind alpha2-macroglobulin (alpha2M) and show homology to
CC protein G of group G Streptococcus. GRAB protein and peptides derived
CC from it are used in vaccine compositions for generating a protective
CC immune response against group A Streptococcus. Antibodies against GRAB
CC are useful for treating Streptococcus pyogenes infections. The protein
CC is also useful for purifying alpha2M from a sample. The present sequence
CC is a DNA encoding full-length GRAB protein from S. pyogenes strain SF370.
XX
SQ Sequence 764 BP; 279 A; 131 C; 159 G; 195 T; 0 other;
XX
Alignment Scores:
Pred. No.: 8,27e-13 Length: 764
Score: 129.00 Matches: 28
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Gaps: 0
DB: 21
XX
US-09-847-539a-6_COPY_59_86 (1-28) x AAD00559 (1-764)
QY 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluAlaAla 20
Db 333 TCAGATGCGCTTAGAAGCATTTAGCGGATCAACAGACGCTTACATCAATGAGAGAGCTGCG 392
```

OY 21 ValValLyAlaAspAsnAlaAla 28  
|||||  
DB 393 GTTGTAAAGCGATACGCTGCT 416

RESULT 5  
AAD00561

ID AAD00561 standard; DNA; 777 BP.

AC AAD00561;

DT 29-AUG-2000 (first entry)

DE Streptococcus pyogenes strain KTL9 partial GRAB protein encoding DNA.

KM GRAB protein; protein G related alpha2M binding protein; vaccine;

KW alpha2-macroglobulin; group A Streptococcus; GAS; antidiabetic;

KM Immune response; Streptococcus pyogenes infection; ds.

OS Streptococcus pyogenes.

FH Key Location/Qualifiers

FT CDS 1..777

FT /\*tag= a

FT /product= "GRAB protein"

FT /note= "Does not include stop codon"

FT /partial

PN WO200026240-A2.

PD 11-MAY-2000.

PF 02-NOV-1999; 99WO-GB03631.

PR 02-NOV-1998; 98GB-0023975.

PR (ACTI-) ACTINOVA LTD.

PA Bjorck LH, Rasmussen M;

PI WPI: 2000-365572/31.

DR P-PSDB: AAY71043.

PT New alpha2M binding protein for generating a protective immune response

PT to group A streptococcus and purifying the binding protein -

PS Claim 13; Page 63-64; 67pp; English.

CC The patent discloses a new family of proteins termed GRAB (protein G  
CC related alpha2M binding protein) from Streptococcus pyogenes which have  
CC the ability to bind alpha2-macroglobulin (alpha2M) and show homology to  
CC protein G of group G Streptococcus. GRAB protein and peptides derived  
CC from it are used in vaccine compositions for generating a protective  
CC immune response against group A Streptococcus. Antibodies against GRAB  
CC are useful for treating Streptococcus pyogenes infections. The protein  
CC is also useful for purifying alpha2M from a sample. The present sequence  
CC is a DNA encoding partial GRAB protein from S. pyogenes strain KTL9.  
CC The protein has alpha2M binding region and is useful in vaccine  
CC composition.

SQ Sequence 777 BP; 269 A; 151 C; 178 G; 179 T; 0 other;

Alignment Scores:

Pred. No.: 8.44e-13 Length: 777

Score: 129.00 Matches: 28

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 100.00% Indels: 0

DB: 21 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x AAD00561 (1-777)

DB 319 AGTAGCGCTTAGACATGTGGCGATCAAGACGAGCTTTACATCAGACAGAGCTGGC 378  
|||||

OY 21 ValValLyAlaAspAsnAlaAla 28  
|||||

DB 379 GTTGTAAAGCGATACGCTGCT 402

RESULT 6  
AAD00563

ID AAD00563 standard; DNA; 853 BP.

AC AAD00563;

DT 29-AUG-2000 (first entry)

DE Streptococcus pyogenes strain AP49 partial GRAB protein encoding DNA.

KM GRAB protein; protein G related alpha2M binding protein; vaccine;

KW alpha2-macroglobulin; group A Streptococcus; GAS; antidiabetic;

KM Immune response; Streptococcus pyogenes infection; ds.

OS Streptococcus pyogenes.

FH Key Location/Qualifiers

FT CDS 37..852

FT /\*tag= a

FT /product= "GRAB protein"

FT /partial

PN WO200026240-A2.

PD 11-MAY-2000.

PF 02-NOV-1999; 99WO-GB03631.

PR 02-NOV-1998; 98GB-0023975.

PR (ACTI-) ACTINOVA LTD.

PA Bjorck LH, Rasmussen M;

PI WPI: 2000-365572/31.

DR P-PSDB: AAY71045.

PT New alpha2M binding protein for generating a protective immune response

PT to group A streptococcus and purifying the binding protein -

PS Claim 13; Page 64; 67pp; English.

CC The patent discloses a new family of proteins termed GRAB (protein G  
CC related alpha2M binding protein) from Streptococcus pyogenes which have  
CC the ability to bind alpha2-macroglobulin (alpha2M) and show homology to  
CC protein G of group G Streptococcus. GRAB protein and peptides derived  
CC from it are used in vaccine compositions for generating a protective  
CC immune response against group A Streptococcus. Antibodies against GRAB  
CC are useful for treating Streptococcus pyogenes infections. The protein  
CC is also useful for purifying alpha2M from a sample. The present sequence  
CC is a DNA encoding partial GRAB protein from S. pyogenes strain AP49.  
CC The protein has alpha2M binding region and is useful in vaccine  
CC composition.

SQ Sequence 853 BP; 295 A; 171 C; 197 G; 190 T; 0 other;

Alignment Scores:

Pred. No.: 9.5e-13 Length: 853

Score: 129.00 Matches: 28

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 100.00% Indels: 0

DB: 21 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x AAD00563 (1-853)



```

OY      1 SerAspAlaLeuGluaAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAla 20
DB      556 AGTACGCCCTTAGACGATTGGCGGATCAACAGACGCTTACATCAGAAAGAGCTGCG 615
OY      21 ValValIysAlaAspAsnAlaAla 28
DB      616 GTTGTAAACGGATACGCTGCT 639
RESULT 7
ID      AAT67618 standard; DNA: 576 BP.
XX
AC      AAT67618:
XX
DT      10-JUL-1997 (first entry)
DE      H. pylori secreted or periplasmic protein ORF 35336707.aa.
XX
KM      Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
KM      identification; binding compound; bacterium; life cycle; activator;
KM      bacteria; inhibitor; duodenal ulcer disease; chronic gastritis;
KM      diagnosis; ds.
XX
OS      Helicobacter pylori.
XX
FH      Key
FT      CDS
FT      1..576
FT      /*tag= a
FT      /transl_except= (pos: 541..543, aa: Glu)
FT      /transl_except= (pos: 547..549, aa: Asp)
FT      /note= "no stop codon given"
FT      559..561
FT      /*tag= b
FT      /note= "encodes Asn"
FT      574..576
FT      /*tag= c
FT      /note= "encodes Asn"
XX
PN      W09640893-A1.
XX
PD      19-DEC-1996.
XX
PF      06-JUN-1996; 96MO-US09122.
XX
PR      01-APR-1996; 96US-0630405.
PR      07-JUN-1995; 95US-0487032.
XX
PA      (ASTR ) ASTRA AB.
PI      Berglindh OT, Smith D, Mellgaerd BL;
XX
DR      WPI; 1997-052306/05.
DR      P-PSDB; AAW20445.
XX
PT      Helicobacter pylori nucleic acid sequences and related
PT      polypeptide(s) - useful for vaccines to treat or prevent H. pylori
PT      infection, and to detect Helicobacter
XX
PS      Claim 23; Page 268; 1481bp; English.
XX
CC      This sequence encodes a H. pylori secreted or periplasmic protein.
CC      The protein may be used in a vaccine to prevent or treat H. pylori
CC      infection or to identify H. pylori polypeptide binding compounds,
CC      useful as potential H. pylori life cycle activators or inhibitors.
CC      The genomic sequence of H. pylori (ATCC 55679) was determined from
CC      overlapping contigs generated by mechanically shearing the bacterial
CC      DNA. The sequences were analysed for ORF of at least 180 nucleotides,
CC      and the predicted coding regions defined by computer evaluation. To
CC      identify likely H. pylori antigens for vaccine development, the amino
CC      acid sequences predicted from various ORF were analysed for significant
CC      homology to other known or exported membrane proteins. Having identified
CC      and determined the sequences of interest, particular regions can be
CC      isolated from H. pylori by PCR amplification for recombinant polypeptide

```

```

CC      production, e.g. in E. coli hosts.
XX
SQ      Sequence 576 BP: 217 A; 97 C; 95 G; 161 T; 6 other:
XX
Alignment Scores:
Pred. No.: 1.02 Length: 576
Score: 57.00 Matches: 12
Percent Similarity: 69.57% Conservative: 4
Best Local Similarity: 52.17% Mismatches: 7
Query Match: 44.19% Indels: 0
DB: 18 Gaps: 0
US-09-847-539a-6_COPY_59_86 (1-28) x AAT67618 (1-576)
OY      4 LeuGluaAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaValValIys 23
DB      145 ATACACGCCCTTAGACGATCAACAAATTGACGCTTACATCAGAAAGAGCTTAGCACA 204
OY      24 AlaAspAsn 26
DB      205 TGGGATTAAC 213
RESULT 8
ID      AAT67891 standard; DNA: 666 BP.
XX
AC      AAT67891:
XX
DT      14-JUL-1997 (first entry)
DE      H. pylori secreted or periplasmic protein ORF 02ce10216orfl.
XX
KM      Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
KM      identification; binding compound; bacterium; life cycle; activator;
KM      bacteria; inhibitor; duodenal ulcer disease; chronic gastritis;
KM      diagnosis; ds.
XX
OS      Helicobacter pylori.
XX
FH      Key
FT      CDS
FT      1..666
FT      /*tag= a
FT      /note= "no stop codon given"
XX
PN      W09640893-A1.
XX
PD      19-DEC-1996.
XX
PF      06-JUN-1996; 96MO-US09122.
XX
PR      01-APR-1996; 96US-0630405.
PR      07-JUN-1995; 95US-0487032.
XX
PA      (ASTR ) ASTRA AB.
PI      Berglindh OT, Smith D, Mellgaerd BL;
XX
DR      WPI; 1997-052306/05.
DR      P-PSDB; AAW20638.
XX
PT      Helicobacter pylori nucleic acid sequences and related
PT      polypeptide(s) - useful for vaccines to treat or prevent H. pylori
PT      infection, and to detect Helicobacter
XX
PS      Claim 23; Page 770-771; 1481bp; English.
XX
CC      The present sequence encodes a H. pylori secreted or periplasmic protein.
CC      The protein may be used in a vaccine to prevent or treat H. pylori
CC      infection or to identify H. pylori polypeptide binding compounds,
CC      useful as potential H. pylori life cycle activators or inhibitors.
CC      The genomic sequence of H. pylori (ATCC 55679) was determined from
CC      overlapping contigs generated by mechanically shearing the bacterial
CC      DNA. The sequences were analysed for ORF of at least 180 nucleotides,

```

CC and the predicted coding regions defined by computer evaluation. To  
CC identify likely H. pylori antigens for vaccine development, the amino  
CC acid sequences predicted from various ORF were analysed for significant  
CC homology to other known or exported membrane proteins. Having identified  
CC and determined the sequences of interest, particular regions can be  
CC isolated from H. pylori by PCR amplification for recombinant polypeptide  
CC production, e.g. in E. coli hosts.

XX  
SQ Sequence 666 BP; 250 A; 115 C; 119 G; 182 T; 0 other;

Alignment Scores:  
Pred. No.: 1,23 Length: 666  
Score: 57.00 Matches: 12  
Percent Similarity: 69.57% Conservative: 4  
Best Local Similarity: 52.17% Mismatches: 7  
Query Match: 44.19% Indels: 0  
DB: 18 Gaps: 0

US-09-847-539a-6\_COPY\_59\_86 (1-28) x AAT67891 (1-666)

OY 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValys 23  
DB 229 ATACAAGCCCTACAGAGCAAAATGACGCTTACATTCACAGAAAAGTCGTACCAA 288

OY 24 AlaAspAsn 26  
DB 289 TCGCATAAC 297

RESULT 9  
AAK30461  
ID AAK30461 standard; DNA: 1239 BP.  
XX  
AC AAK30461;  
XX  
DT 08-JUN-1999 (first entry)  
XX  
DE H. pylori secreted protein ORF 09ce10413\_35336707\_f2.9.  
XX  
KW Vaccine: probe: diagnostic; ORF; cell envelope protein;  
XX secreted protein; cellular protein; ds.  
XX  
OS Helicobacter pylori.  
XX  
PN W09818323-A1.  
XX  
PD 07-MAY-1998.  
XX  
PF 28-OCT-1997; 97WO-US19575.  
XX  
PR 14-JUL-1997; 97US-0891928.  
XX 28-OCT-1996; 96US-0739150.  
XX 06-DEC-1996; 96US-0759739.  
XX  
PA (ASTR ) ASTRA AB.  
XX  
PI Alm RA, Smith D;  
XX  
XX WPI: 1998-271811/24.  
XX P-PSDB: AAY10994.  
XX  
XX Helicobacter pylori nucleic acids and proteins - used to develop  
XX products for the detection, prevention and treatment of H. pylori  
XX infections  
XX  
XX Claims 3, 4; Page 126-127; 279pp; English.

CC Recombinant or substantially pure preparations of H. pylori polypeptides  
CC are disclosed, together with the nucleic acids encoding them. In all,  
CC 73 ORFs are shown. The proteins are variously cell envelope proteins,  
CC secreted proteins or other cellular proteins. Vaccines containing the  
CC nucleic acids or proteins are claimed, as are probes containing at least  
CC 8 nucleotides from the nucleic acid sequences. The vaccines are useful  
CC for treating or reducing the risk of H. pylori infections, and the

CC probes can be used diagnostically for detecting the presence of  
CC Helicobacter in a sample. The products are also of use in screening  
CC for compounds having the ability to interfere with the H. pylori life  
CC cycle or to inhibit H. pylori infection.

XX  
SQ Sequence 1239 BP; 472 A; 234 C; 206 G; 327 T; 0 other;

Alignment Scores:  
Pred. No.: 2,69 Length: 1239  
Score: 57.00 Matches: 12  
Percent Similarity: 69.57% Conservative: 4  
Best Local Similarity: 52.17% Mismatches: 7  
Query Match: 44.19% Indels: 0  
DB: 19 Gaps: 0

US-09-847-539a-6\_COPY\_59\_86 (1-28) x AAK30461 (1-1239)

OY 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValys 23  
DB 145 ATACAAGCCCTACAGAGCAAAATGACGCTTACATTCACAGAAAAGTCGTACCAA 204

OY 24 AlaAspAsn 26  
DB 205 TCGCATAAC 213

RESULT 10  
ABL13527/C  
ID ABL13527 standard; CDNA: 3349 BP.  
XX  
AC ABL13527;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 35063.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
XX pharmaceutical; gene; ss.  
XX  
OS Drosophila melanogaster.  
XX  
PN W0200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US09231.  
XX  
PR 23-MAR-2000; 2000US-191637P.  
XX 11-JUL-2000; 2000US-0614150.  
XX  
PA (PEKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li PWD, Myers EW;  
XX  
XX WPI: 2001-656860/75.  
XX P-PSDB: ABB69424.  
XX  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
XX genes from Drosophila and for elucidating cell signalling and cell-cell  
XX interactions -  
XX  
XX Claim 1; SEQ ID NO 35063; 21pp + Sequence listing; English.

CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (AB101840-AB16175) and the encoded proteins  
CC (AB157737-AB12072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pcl\_sequences.

XX SQ Sequence 3349 BP; 808 A; 953 C; 851 G; 737 T; 0 other;  
Alignment Scores:  
Pred. No.: 30.6 Length: 3349  
Score: 54.00 Matches: 12  
Percent Similarity: 66.67% Conservative: 4  
Best Local Similarity: 50.00% Mismatches: 8  
Query Match: 41.86% Indels: 0  
DB: 23 Gaps: 0  
US-09-847-539a-6\_COPY\_59\_86 (1-28) x ABL13527 (1-3349)  
QY 5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValIysAla 24  
DB 1122 GAGGAGGAAGCTGACGAGGAGCGCCGCGAGCGAGATGCCGCGCTGCCGCC 1063  
QY 25 AspasnAlaAla 28  
DB 1062 GATGCTGGCGCG 1051  
RESULT 11  
ABL13526/c  
ID ABL13526 standard; CDNA; 6398 BP.  
XX ABL13526;  
AC  
XX 26-MAR-2002 (first entry)  
DT  
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 35060.  
DE  
XX Drosophila: developmental biology; cell signalling; insecticide;  
KW pharmaceutical; gene; ss.  
XX  
XX Drosophila melanogaster.  
OS  
XX  
XX W0200171042-A2.  
PN  
XX  
PD 27-SEP-2001.  
XX  
XX 23-MAR-2001; 2001WO-US09231.  
PE  
XX 23-MAR-2000; 2000US-191637P.  
PR 11-JUL-2000; 2000US-0614150.  
PR  
XX (PEKE ) PE CORP NY.  
PA  
XX  
XX Venter JC, Adams M, Li PWD, Myers EW;  
PI  
XX WPI: 2001-656860/75.  
DR P-FSDB; ABB69423.  
DR  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
genes from Drosophila and for elucidating cell signalling and cell-cell  
interactions -  
PT  
XX  
XX Claim 1; SEQ ID NO 35060; 21pp + Sequence Listing; English.  
PS  
XX  
XX The invention relates to an isolated nucleic acid detection reagent  
capable of detecting 1000 or more genes from Drosophila. The invention is  
useful in developmental biology and in elucidating cell signalling and  
cell-cell interactions in higher eukaryotes for the development of  
insecticides, therapeutics and pharmaceutical drugs. The invention  
discloses genomic DNA sequences (ABL16176-ABL10511), expressed DNA  
sequences (ABB57737-ABB72072),  
CC (ABB57737-ABB72072).  
CC The sequence data for this patent did not form part of the printed  
specification, but was obtained in electronic format directly from WIPO  
at ftp.wipo.int/pub/published\_pat\_sequences.  
CC  
XX  
XX Sequence 6398 BP; 1688 A; 1632 C; 1352 G; 1726 T; 0 other;  
SQ  
Alignment Scores:

Pred. No.: 69.3 Length: 6398  
Score: 54.00 Matches: 12  
Percent Similarity: 66.67% Conservative: 4  
Best Local Similarity: 50.00% Mismatches: 8  
Query Match: 41.86% Indels: 0  
DB: 23 Gaps: 0  
US-09-847-539a-6\_COPY\_59\_86 (1-28) x ABL13526 (1-6398)  
QY 5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValIysAla 24  
DB 2298 GAGGAGGAAGCTGACGAGGAGCGCCGCGAGCGAGATGCCGCGCTGCCGCC 2239  
QY 25 AspasnAlaAla 28  
DB 2238 GATGCTGGCGCG 2227  
RESULT 12  
AAC54228/c  
ID AAC54228 standard; DNA; 1002 BP.  
XX AAC54228;  
AC  
XX 18-OCT-2000 (first entry)  
DT  
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 77144.  
DE  
XX Hybridisation assay; genetic mapping; gene expression control;  
KW protein identification; signal transduction pathway;  
KW metabolic pathway; promoter; termination sequence; ss.  
XX  
XX Arabidopsis thaliana.  
OS  
XX  
XX EP1033405-A2.  
PN  
XX  
PD 06-SEP-2000.  
XX  
XX 25-FEB-2000; 2000EP-0301439.  
PE  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132863.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 18-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.

PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137528.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139464.  
PR 18-JUN-1999; 99US-0139465.  
PR 18-JUN-1999; 99US-0139466.  
PR 18-JUN-1999; 99US-0139467.  
PR 21-JUN-1999; 99US-0139468.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.

PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151348.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.



PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0148368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0158993.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 23-OCT-1999; 99US-0161404.  
PR 23-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161921.  
PR 28-OCT-1999; 99US-0161922.

PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

## Alignment Scores:

Pred. No.:	18.9	Length:	1230
Score:	52.00	Matches:	11
Percent Similarity:	72.73%	Conservative:	5
Best Local Similarity:	50.00%	Mismatches:	6
Query Match:	40.31%	Indels:	0
DB:	21	Gaps:	0

US-09-847-539a-6\_COPY\_59\_86 (1-28) x AAC33266 (1-1230)

Oy 4 LeuGLuAlaLeuAlaAspClnThrAspAlaLeuGlnSerGlnGluAlaAlaValLys 23

Db 518 CTTCCCGCACACTCTCCGACAAAGGACATGCCCTCAATTCAAAAGCTGCATCGCAAG 459

Oy 24 AlaAsp 25

Db 458 GCGGAT 453

## RESULT 14

ABL25211/c  
ID ABL25211 standard; DNA; 16135 BP.

XX ABL25211;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 27106.

XX Drosophila: developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ds.

XX Drosophila melanogaster.

XX WO200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PEKE ) PE CORP NY.

XX Venter JC, Adams M, Li PMD, Myers EW;

DR WPI, 2001-656860/75.

XX New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from Drosophila and for elucidating cell signalling and cell-cell

PT interactions -

PS Claim 1; SEQ ID NO 27106; 21pp + sequence listing; English.

XX The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and

CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA

CC sequences (AB101840-AB16175) and the encoded proteins

CC (AB57737-AB872072).

CC The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 16135 BP; 4290 A; 3936 C; 4034 G; 3875 T; 0 other;

Alignment Scores: 488 Length: 16135

Pred. No.: 16135

```

Score:          52.00          Matches:          11
Percent Similarity: 65.22%      Conservative: 4
Best Local Similarity: 47.83%    Mismatches: 8
Query Match:      40.31%        Indels: 0
DB:               23           Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x ABL25210 (1-16135)

OY      4 LeuGlUAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValLys 23
         |||||
Db       5971 CTGGAAGCTCTGTTCAGAGCTGAGTCTCTCCAAAGCCGCGAGTTGCTCTTGCGCA 5912
         |||||
OY      24 AlaAspAsn 26
         |||||
Db       5911 GCCGACAAAT 5903

RESULT 15
ABL25210
ID       ABL25210 standard; DNA: 18603 BP.
XX
AC       ABL25210;
XX
DE       26-MAR-2002 (first entry)
XX
DE       Drosophila melanogaster genomic polynucleotide SEQ ID NO 27103.
XX
KM       Drosophila; developmental biology; cell signalling; insecticide;
XX
KW       pharmaceutical; gene; ds.
XX
OS       Drosophila melanogaster.
XX
PN       WO200171042-A2.
XX
PD       27-SEP-2001.
XX
PF       23-MAR-2001; 2001WO-US09231.
XX
PR       23-MAR-2000; 2000US-191637P.
XX
PR       11-JUL-2000; 2000US-0614150.
XX
PA       (PEKE ) PE CORP NY.
XX
PI       Venter JC, Adams M, Li PMD, Myers EM;
XX
DR       WPI; 2001-656860/75.
XX
PT       New isolated nucleic acid detection reagent for detecting 1000 or more
PT       genes from Drosophila and for elucidating cell signalling and cell-cell
PT       interactions -
XX
PS       Claim 1; SEQ ID NO 27103; 21pp + Sequence Listing; English.
XX
CC       The invention relates to an isolated nucleic acid detection reagent
CC       capable of detecting 1000 or more genes from Drosophila. The invention is
CC       useful in developmental biology and in elucidating cell signalling and
CC       cell-cell interactions in higher eukaryotes for the development of
CC       insecticides, therapeutics and pharmaceutical drugs. The invention
CC       discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC       sequences (ABL01840-ABL16175) and the encoded proteins
CC       (ABBS7737-ABR72072).
CC       The sequence data for this patent did not form part of the printed
CC       specification, but was obtained in electronic format directly from WIPO
CC       at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ       Sequence 18603 BP; 4763 A; 4409 C; 4291 G; 5140 T; 0 other;

Alignment Scores:
Pred. No.:      584          Length:          18603
Score:          52.00          Matches:          11
Percent Similarity: 65.22%      Conservative: 4
Best Local Similarity: 47.83%    Mismatches: 8
Query Match:      40.31%        Indels: 0
DB:             23           Gaps: 0

```

```

US-09-847-539A-6_COPY_59_86 (1-28) x ABL25210 (1-18603)

OY      4 LeuGlUAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValLys 23
         |||||
Db       11563 CTGGAAGCTCTGTTCAGAGCTGAGTCTCTCCAAAGCCGCGAGTTGCTCTTGCGCA 11622
         |||||
OY      24 AlaAspAsn 26
         |||||
Db       11623 GCCGACAAAT 11631

RESULT 16
ABL01876/c
ID       ABL01876 standard; cDNA: 25461 BP.
XX
AC       ABL01876;
XX
DE       26-MAR-2002 (first entry)
XX
DE       Drosophila melanogaster expressed polynucleotide SEQ ID NO 110.
XX
KM       Drosophila; developmental biology; cell signalling; insecticide;
XX
KW       pharmaceutical; gene; ss.
XX
OS       Drosophila melanogaster.
XX
PN       WO200171042-A2.
XX
PD       27-SEP-2001.
XX
PF       23-MAR-2001; 2001WO-US09231.
XX
PR       23-MAR-2000; 2000US-191637P.
XX
PR       11-JUL-2000; 2000US-0614150.
XX
PA       (PEKE ) PE CORP NY.
XX
PI       Venter JC, Adams M, Li PMD, Myers EM;
XX
DR       WPI; 2001-656860/75.
XX
DR       P-PSDB; ABB57773.
XX
PT       New isolated nucleic acid detection reagent for detecting 1000 or more
PT       genes from Drosophila and for elucidating cell signalling and cell-cell
PT       interactions -
XX
PS       Claim 1; SEQ ID NO 110; 21pp + Sequence Listing; English.
XX
CC       The invention relates to an isolated nucleic acid detection reagent
CC       capable of detecting 1000 or more genes from Drosophila. The invention is
CC       useful in developmental biology and in elucidating cell signalling and
CC       cell-cell interactions in higher eukaryotes for the development of
CC       insecticides, therapeutics and pharmaceutical drugs. The invention
CC       discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC       sequences (ABL01840-ABL16175) and the encoded proteins
CC       (ABBS7737-ABR72072).
CC       The sequence data for this patent did not form part of the printed
CC       specification, but was obtained in electronic format directly from WIPO
CC       at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ       Sequence 25461 BP; 7542 A; 5394 C; 5424 G; 7101 T; 0 other;

Alignment Scores:
Pred. No.:      1,06e+03       Length:          25461
Score:          51.50          Matches:          13
Percent Similarity: 65.38%      Conservative: 4
Best Local Similarity: 50.00%    Mismatches: 8
Query Match:      39.92%        Indels: 1
DB:             23           Gaps: 1

US-09-847-539A-6_COPY_59_86 (1-28) x ABL01876 (1-25461)

OY      2 AspAlaLeuGlUAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaVal 21

```

```

Db 12848 GACGCTTTAAAGCGCTTTCTTCACAGACTGAT---TTGAGCCATCCGATGAGTGTATTT 12792
OY 22 VallysAlaAspAsnAla 27
Db 12791 TTTAAGCAGACCAATGCC 12774

RESULT 17
ABL12683/C
ID ABL12683 standard; cDNA; 1960 BP.
XX
AC ABL12683;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 32531.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
WP1: 2001-656860/75.
DR P-PSDB; ABB68580.
XX
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 32531; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB16176-ABL30511), expressed DNA
CC sequences (AB101840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcr_sequences.
XX
SQ Sequence 1960 BP; 505 A; 526 C; 485 G; 444 T; 0 other;

Alignment Scores:
Pred. No.: 50.3 Length: 1960
Score: 51.00 Matches: 11
Percent Similarity: 53.85% Conservative: 3
Best Local Similarity: 42.31% Mismatches: 12
Query Match: 39.53% Indels: 0
DB: 23 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x ABL12683 (1-1960)
OY 2 AspaAlaLeuGluAlaLeuAlaAspGlnThrAspaAlaLeuGlnSerGluGluAlaAlaVal 21
Db 228 GATCGCGGCAATCATATGAGAACACACAGACGAGAGAGACGAGAGAACACAGCGCGC 169
OY 22 VallysAlaAspAsnAla 27

```

```

Db 168 GTACGAGAGATCGTGC 151
OY 22 VallysAlaAspAsnAla 27
Db 168 GTACGAGAGATCGTGC 151

RESULT 18
ABL27360
ID ABL27360 standard; DNA; 2618 BP.
XX
AC ABL27360;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 33553.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
WP1: 2001-656860/75.
XX
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 33553; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB16176-ABL30511), expressed DNA
CC sequences (AB101840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcr_sequences.
XX
SQ Sequence 2618 BP; 669 A; 603 C; 652 G; 694 T; 0 other;

Alignment Scores:
Pred. No.: 72.5 Length: 2618
Score: 51.00 Matches: 11
Percent Similarity: 53.85% Conservative: 3
Best Local Similarity: 42.31% Mismatches: 12
Query Match: 39.53% Indels: 0
DB: 23 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x ABL27360 (1-2618)
OY 2 AspaAlaLeuGluAlaLeuAlaAspGlnThrAspaAlaLeuGlnSerGluGluAlaAlaVal 21
Db 467 GATCGCGGCAATCATATGAGAACACACAGACGAGAGAGACGAGAGAACACAGCGCGC 526
OY 22 VallysAlaAspAsnAla 27
Db 527 GTACGAGAGATCGTGC 544
OY 22 VallysAlaAspAsnAla 27

```



ID ABL12682 standard; cDNA; 4221 BP.  
XX  
AC ABL12682:  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster expressed polynucleotide seq ID NO 32528.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KM pharmaceutical; gene; ss.  
XX  
OS Drosophila melanogaster.  
XX  
PN WO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US09231.  
XX  
PR 23-MAR-2000; 2000US-191637P.  
PR 11-JUL-2000; 2000US-0614150.  
XX  
PA (PEKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li PMD, Myers EM;  
PI WPI; 2001-656860/75.  
DR F-PSDB; ABB68579.  
XX  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX  
PS Claim 1: SEQ ID NO 32528; 21bp + Sequence Listing; English.  
XX  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA  
CC sequences (AB101840-AB16175) and the encoded proteins  
CC (AB57737-AB572072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 4221 BP; 1022 A; 1036 C; 1072 G; 1091 T; 0 other;

Alignment Scores:  
Pred. No.: 133 Length: 4221  
Score: 51.00 Matches: 11  
Percent Similarity: 53.858 Conservative: 3  
Best Local Similarity: 42.318 Mismatches: 12  
Query Match: 39.538 Indels: 0  
DB: 23 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x ABL12682 (1-4221)  
QY 2 AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuInserGluAlaAlaVal 21  
DB 2994 GATCGGCGGAAATCATATAGAAACACACAGACGACGAGAGAGAGAGAGAAACAAACGCGCG 3053  
QY 22 ValValysAlaAspAsnAla 27  
DB 3054 GTACGAGAGATCGTCC 3071

RESULT 20  
ABAO3041  
ID ABA03041 standard; DNA; 2944528 BP.  
XX  
AC ABA03041:  
XX

DT 05-FEB-2002 (first entry)  
XX  
XX  
DE Listeria monocytogenes EGD-e genome sequence.  
XX  
KW Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;  
KM vitamin B12; bacterial infection; disease; ds.  
XX  
OS Listeria monocytogenes.  
XX  
PN WO200177335-A2.  
XX  
PD 18-OCT-2001.  
XX  
PF 11-APR-2001; 2001WO-FR01118.  
XX  
PR 11-APR-2000; 2000FR-0004629.  
XX  
PA (INSP ) INST PASTEUR.  
XX  
PI Buchrieser C, Frangeul L, Couve E, Rusniok C, Psini H, Dehoux P;  
PI Dussurget O, Chetoui F, Nedjati H, Glaser P, Kunst F, Cossart P;  
PI Daniels J, Goebel W, Kreft J, Kuhn M, Ng E, Vazquez-Boland JA;  
PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;  
PI Chakraborty T, Domann E, Hain T, Berche P, Chardit A, Durant L;  
PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;  
PI Madueno E, De Pablo B, Weiland J, Kaerst U, Entian K, Haut J;  
PI Rose M, Voss H;  
DR WPI; 2002-010914/01.  
XX  
XX  
PT Genomic sequence for Listeria monocytogenes, useful e.g. for treatment  
PT and prevention of Listeria and related bacterial infections, and  
PT related polypeptides -  
XX  
PS Claim 1: SEQ ID NO 1; 192pp; French.  
XX  
XX  
CC The present sequence is the genome sequence of Listeria monocytogenes  
CC EGD-e. This sequence and fragments of this sequence are useful for  
CC selecting probes and primers for detecting genes in L. monocytogenes and  
CC related organisms, and to study genetic polymorphisms and other genomes.  
CC proteins (AB547297-AB550149) expressed from the present sequence are  
CC useful for raising specific antibodies, identification of L.  
CC monocytogenes and related organisms, and for biosynthesis and  
CC biodegradation, especially biosynthesis of vitamin B12. This sequence and  
CC proteins encoded by it are also useful for selecting compounds that  
CC regulate gene expression and cell replication and modulate L.  
CC monocytogenes-related diseases. In addition, this sequence and proteins  
CC encoded by it are useful in pharmaceutical and vaccine compositions for  
CC the treatment or prevention of infections by L. monocytogenes and related  
CC organisms.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 2944528 BP; 914202 A; 563301 C; 555061 G; 911964 T; 0 other;

Alignment Scores:  
Pred. No.: 4.83e+05 Length: 2944528  
Score: 51.00 Matches: 10  
Percent Similarity: 65.388 Conservative: 7  
Best Local Similarity: 38.464 Mismatches: 9  
Query Match: 39.538 Indels: 0  
DB: 24 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x ABA03041 (1-2944528)  
QY 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuInserGluAlaAla 20  
DB 51231 GCTGAACCAAAAGAACACACTGAAGTGAACACGATCGACGAGAGATGACGT 51290  
QY 21 ValValysAlaAspAsn 26  
DB 51291 GGAAGTGAACAAAGAAAC 51308





APPLICATION NUMBER: US/08/591,079  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Livnat, Shmuel  
REGISTRATION NUMBER: 33,949  
REFERENCE/DOCKET NUMBER: 15661-20017.00  
TELEPHONE: (202) 887-1500  
TELEFAX: (202) 887-0764  
TELEX: 90-4030 MRSNFOERSWSH  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5393 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORGANISM: Salmonella typhi  
STRAIN: Ty2  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 543..2324  
OTHER INFORMATION: /gene= "sfpB"  
US-08-591-079-9

Alignment Scores:  
Pred. No.: 13.3 Length: 5393  
Score: 66.00 Matches: 23  
Percent Similarity: 40.70% Conservative: 12  
Best Local Similarity: 26.74% Mismatches: 51  
Query Match: 8.43% Indels: 0  
DB: 2 Gaps: 0

US-09-847-539A-6 (1-159) x US-08-591-079-9 (1-5393)

Qy 58 SerSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAla 77  
Db 957 CGGACGGATCTCTATGACCCAGCATCAAAAGAGCGGATACCCCAAGAGTGTATGAC 1016  
Qy 78 AlaValValLysAlaAspAsnAlaAlaSerAspAlaLeuGluAlaLeuAlaAspGlnThr 97  
Db 1017 GCTGGCGGCAAAAACTGACGCGAGCGCAAAATAATTCATCGCTGACCCAGCTGAC 1076  
Qy 98 AspAlaLeuGlnSerGluGluAlaGluValGlnSerAspAsnAlaAlaSerAspAla 117  
Db 1077 CCGCGCTATGCACAAGCTCAAGCCCGCTAGAACACAGCGCGGAAAAAGACGAGAGCGG 1136  
Qy 118 TtpGluLysAlaAlaThrProfileAlaLeuAspValLysLysThrLysAspThrLysPro 137  
Db 1137 AAAGAGCGCTTAGATAAGCCACGATCGGACGGTTAAGCAGGACACAGCCCAAGGG 1196  
Qy 138 ValValLysLysGluGlu 143  
Db 1197 AAAGCCGAGAAAGCGGAT 1214

RESULT 16  
US-08-728-323A-1  
Sequence 1, Application US/08728323A  
Patent No. 5948676  
GENERAL INFORMATION:

APPLICANT: Chang, Yuan  
APPLICANT: Bohenzky, Roy A.  
APPLICANT: Russo, James J.  
APPLICANT: Edelman, Isidore S.  
APPLICANT: Moore, Patrick S.  
TITLE OF INVENTION: Immediate Early Protein From Kaposi's  
TITLE OF INVENTION: Sarcoma-Associated Herpesvirus, DNA  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cooper & Dunham LLP

STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/728,323A  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 0575/52268/JPW/MSC/SKS  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-278-0400  
TELEFAX: 212-391-0525  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3489 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..3489  
US-08-728-323A-1

Alignment Scores:  
Pred. No.: 10.5 Length: 3489  
Score: 65.00 Matches: 18  
Percent Similarity: 45.69% Conservative: 35  
Best Local Similarity: 15.52% Mismatches: 63  
Query Match: 8.37% Indels: 0  
DB: 2 Gaps: 0

US-09-847-539A-6 (1-159) x US-08-728-323A-1 (1-3489)

Qy 30 LeuArgAsnGluGluArgAlaIleAspGluLeuLysLysGlnAlaIleGluAspLysGlu 49  
Db 2326 TTAGAGGATCAGGAGGAGGAGTTAGAGGAGCAGGAGGAGGAGGAGGAGGAGGAG 2385  
Qy 50 AlaThrThrAlaIleGluAlaAlaSerSerAspAlaLeuGluAlaLeuAlaAspGlnThr 69  
Db 2386 GAGTTAGAGGAGCAGGAGGAGGAGTTAGAGGAGCAGGAGGAGGAGGAGGAGGAG 2445  
Qy 70 AspAlaLeuGlnSerGluGluAlaAlaValValLysAlaAspAsnAlaAlaSerAspAla 89  
Db 2446 CAGGAGTTAGAGGAGCAGGAGGAGTTAGAGGAGCAGGAGGAGGAGGAGGAGGAG 2505  
Qy 90 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaGluValValGln 109  
Db 2506 GAGCAGGAGTTAGAGGAGCAGGAGGAGTTAGAGGAGCAGGAGGAGGAGGAGGAG 2565  
Qy 110 SerAspAsnAlaAlaSerAspAlaTrpGluLysAlaAlaThrProfileAlaLeuAspVal 129  
Db 2566 CAGGAGCAGGAGCAGGAGGAGTTAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2625  
Qy 130 LysLysThrLysAspThrLysProValValLysLysLysGluGluArgGln 145  
Db 2626 CAGGAGGAGCAGGAGGAGTTAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2673

RESULT 17  
US-08-770-379-20/c  
Sequence 20, Application US/08770379  
Patent No. 5849564  
GENERAL INFORMATION:  
APPLICANT: Chang, Yuan

```
;
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1419 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1416
; US-09-115-746-9
;
; Alignment Scores:
; Pred. No.: 1.57 Length: 1419
; Score: 67.00 Matches: 18
; Percent Similarity: 42.22% Conservative: 20
; Best Local Similarity: 20.00% Mismatches: 52
; Query Match: 8.62% Indels: 0
; DB: 4 Gaps: 0
;
; US-09-847-539a-6 (1-159) x US-09-115-746-9 (1-1419)
;
; Qy 33 GluGluArgAlaIleAspGluLeuLysGlnAlaIleGluAspLysGluAlaThrThr 52
; |||:||||| ||| ||| |||:|||||
; Db 763 GAGAACGAGCGGCGAGCTGAAGCCACGAGGTTGCCGAAGCGGAGACGACGAGCT 822
;
; Qy 53 AlaIleGluAlaAlaSerSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeu 72
; |||: ||| ||| |||: |||
; Db 823 GAAGCCACGAGGTTGCCGAAGCGGAGAGGAGCGGCGAGCTGAAGCCACGAGGTTGCC 882
;
; Qy 73 GlnSerGluGluAlaAlaValValLysAlaAspAsnAlaAlaSerAspAlaLeuGluAla 92
; |||:||||| ||| ||| |||: |||
; Db 883 GAAGCGGAGAGCAAGGCGAGCTGAGCCACGAGGTTGCCGAGGAGCGAGGAGGAGAG 942
;
; Qy 93 LeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaGluValValGlnSerAspAsn 112
; |||: ||| |||: ||| |||: |||
; Db 943 GCAGCTGAAGCCACGAGGTTGCCGAAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1002
;
; Qy 113 AlaAlaSerAspAlaTrpGluLysAlaAla 122
; ||| ||| |||: ||| ||| |||
; Db 1003 GTTCCGAAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGCT 1032
;
; RESULT 14
; US-09-591-079-7
; Sequence 7, Application US/08591079
; Patent No. 5972899
; GENERAL INFORMATION:
; APPLICANT: Zychlinsky, Arturo
; APPLICANT: Chen, Yajing
; TITLE OF INVENTION: Apoptosis Induced by Shigella ipab
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 Pennsylvania Avenue, NW
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20006-1812
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/591.079
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Livnat, Shmuel
; REGISTRATION NUMBER: 33,949
;
; REFERENCE/DOCKET NUMBER: 15661-20017.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-1500
; TELEFAX: (202) 887-0764
; TELEX: 90-4030 MRSNFOERSWSH
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3622 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; ORIGINAL SOURCE:
; ORGANISM: Salmonella typhimurium
; STRAIN: sibB
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 575..2356
; OTHER INFORMATION: /product= "sipB"
; OTHER INFORMATION: /gene= "sipB"
; US-08-591-079-7
;
; Alignment Scores:
; Pred. No.: 7.8 Length: 3622
; Score: 66.00 Matches: 23
; Percent Similarity: 40.70% Conservative: 12
; Best Local Similarity: 26.74% Mismatches: 51
; Query Match: 8.49% Indels: 0
; DB: 2 Gaps: 0
;
; US-09-847-539a-6 (1-159) x US-08-591-079-7 (1-3622)
;
; Qy 58 SerSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAla 77
; |||:||||| ||| ||| |||: |||
; Db 989 GCGAGCGATCTATGAAGCCAGTATCAAAAGACGAGTACCGCCACGAGTGTATTATGAC 1048
;
; Qy 78 AlaValValLysAlaAspAsnAlaAlaSerAspAlaLeuGluAlaLeuAlaAspGlnThr 97
; ||| ||| ||| ||| |||: |||
; Db 1049 GCTCGACCAAAACTGACGCGCGCAAAATAAATTGCAATCGCTGGACCCGCGCTGAC 1108
;
; Qy 98 AspAlaLeuGlnSerGluGluAlaGluValValGlnSerAspAsnAlaAlaSerAspAla 117
; ||| ||| ||| ||| |||: |||
; Db 1109 CCGGCTATGCACAGCTGAAGCCGCGGTAGAACGCGGCGGAGGAGGAGGAGGAGGAGG 1168
;
; Qy 118 TrpGluLysAlaAlaThrProIleAlaLeuAspValLysLysThrLysAspThrLysPro 137
; ||| ||| ||| ||| |||
; Db 1169 AAAGAGCGCTTAGATAAGCCACGAGTACGAGCGGTTAAAGCAGGCGACAGCCCAAGCG 1228
;
; Qy 138 ValValLysLysGluGlu 143
; ||| ||| |||
; Db 1229 AAAGCCGAGAAAGCGGAT 1246
;
; RESULT 15
; US-08-591-079-9
; Sequence 9, Application US/08591079
; Patent No. 5972899
; GENERAL INFORMATION:
; APPLICANT: Zychlinsky, Arturo
; APPLICANT: Chen, Yajing
; TITLE OF INVENTION: Apoptosis Induced by Shigella ipab
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 Pennsylvania Avenue, NW
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20006-1812
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/591.079
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Livnat, Shmuel
; REGISTRATION NUMBER: 33,949
;
; ;
```



; TITLE OF INVENTION: No. 6348582el Prokaryotic Polynucleotides,  
; TITLE OF INVENTION: Polypeptides and their Uses  
; NUMBER OF SEQUENCES: 534  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Smithkline Beecham Corporation  
; STREET: 709 Swedeland Road  
; CITY: King of Prussia  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19406-0939  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/936.165A  
; FILING DATE: 24-SEP-1997  
; CLASSIFICATION: 536

; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/027.032  
; FILING DATE: 24-SEP-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Gimmi, Edward R  
; REGISTRATION NUMBER: 38,891  
; REFERENCE/DOCKET NUMBER: P50549  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 610-270-4478  
; TELEFAX: 610-270-5090  
; TELEX:

; INFORMATION FOR SEQ ID NO: 93:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 605 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: Genomic DNA  
US-08-936-165A-93

Alignment Scores:  
Pred. No.: 0.127 Length: 605  
Score: 71.00 Matches: 20  
Percent Similarity: 41.49% Conservative: 19  
Best Local Similarity: 21.28% Mismatches: 55  
Query Match: 9.14% Indels: 0  
DB: 4 Gaps: 0

US-09-847-539A-6 (1-159) x US-08-936-165A-93 (1-605)  
Qy 62 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLys 81  
Db 588 CTCGAGGCATGCAAGCTTGAGTCTTNTAATTCAGACACACACACACACAGT 529  
Qy 82 AlaAspAsnAlaAlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGln 101  
Db 528 AATCGAATGCGAAAATGAAGTGTGATCAAGCTGTGACAACTCAAAATCAAGCAATGAT 469  
Qy 102 SerGluGluAlaGluAlaValValGlnSerAspAsnAlaAlaSerAspAlaTrpGluLysAla 121  
Db 468 AATCACTGGCGCTACAACTGAGAGAAAATGACGAAAAGATTTAGTTTANAAGCT 409  
Qy 122 AlaThrProIleAlaLeuAspValLysLysThrLysAspThrLysProValValLysLys 141  
Db 408 AAGAAAAAGCGTNTCAAGATATCTTAATGTCACAAACAACTAATGATGTTACGCAAT 349  
Qy 142 GluGluArgGlnAsnValAsnThrLeuProThrGlyGlu 155  
Db 348 AATGATCAAGCAGTGTCTGATATTCAGAGTATTCAAGGTATTACTGGAGAT 307

RESULT 10  
US-08-973-462-2  
; Sequence 2, Application US/08973462B  
; Patent No. 6191270

; GENERAL INFORMATION:  
; APPLICANT: DAUBERSIES, PIERRE  
; TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES  
; FILE REFERENCE: 0660-0125-0 PCT  
; CURRENT APPLICATION NUMBER: US/08/973.462B  
; EARLIER FILING DATE: 1998-02-06  
; EARLIER APPLICATION NUMBER: PCT/FR96/00894  
; EARLIER FILING DATE: 1996-06-12  
; EARLIER APPLICATION NUMBER: FR 95/07007  
; EARLIER FILING DATE: 1995-06-13  
; NUMBER OF SEQ ID NOS: 29  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 5361  
; TYPE: DNA  
; ORGANISM: P. falciparum  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(5361)  
US-08-973-462-2

Alignment Scores:  
Pred. No.: 6.66 Length: 5361  
Score: 68.00 Matches: 17  
Percent Similarity: 42.06% Conservative: 28  
Best Local Similarity: 15.89% Mismatches: 62  
Query Match: 8.75% Indels: 0  
DB: 4 Gaps: 0

US-09-847-539A-6 (1-159) x US-08-973-462-2 (1-5361)  
Qy 26 GluLysLeuAlaLeuAlaArgAsnGluGluAlaIleAspGluLeuLysLysGlnAlaIle 45  
Db 682 CAAGAACTGTAGAGAAAATGACCAAGAAAGTGTAGAAGAAAATGTAGAAGAAAATGTA 741  
Qy 46 GluAspLysGluAlaThrThrAlaIleGluAlaAlaSerSerAspAlaLeuGluAlaLeu 65  
Db 742 GAAGAAAATGACGACGGAAGTGTAGCTCAAGTGTGAAGAAAGTATAGCTTCAAGTGT 801  
Qy 66 AlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLysAlaAspAsnAla 85  
Db 802 GATGAAAGTATAGATTCAAGTATTGAAAGAAAATGATGCTTCAACTGTTGAAGAAAATCGTA 861  
Qy 86 AlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAla 105  
Db 862 GCTCAAGTGTGTAGAAAGTGTGCTCCAACTGTTGAAGAAAGTGTAGAAGAAAATGTT 921  
Qy 106 GluValValGlnSerAspAsnAlaAlaSerAspAlaTrpGluLysAlaAlaThrProIle 125  
Db 922 GAAGAAAGTGTAGCTGAAAATGTTGAAGAAAGTGTAGCTGAAAATGTTGAAGAAAATGTA 981  
Qy 126 AlaLeuAspValLysLysThr 132  
Db 982 GCTGAAAATGTTGAAGAAAGT 1002

RESULT 11  
US-08-973-462-1  
; Sequence 1, Application US/08973462B  
; Patent No. 6191270  
; GENERAL INFORMATION:  
; APPLICANT: DAUBERSIES, PIERRE  
; TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES  
; FILE REFERENCE: 0660-0125-0 PCT  
; CURRENT APPLICATION NUMBER: US/08/973.462B  
; EARLIER FILING DATE: 1998-02-06  
; EARLIER APPLICATION NUMBER: PCT/FR96/00894  
; EARLIER FILING DATE: 1996-06-12  
; EARLIER APPLICATION NUMBER: FR 95/07007  
; EARLIER FILING DATE: 1995-06-13  
; NUMBER OF SEQ ID NOS: 29  
; SOFTWARE: PatentIn Ver. 2.0





```
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: German P 42 36 708.5
; FILING DATE: October 30, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Kurt G. Briscoe
; REGISTRATION NUMBER: 33,141
; REFERENCE/DOCKET NUMBER: Bayer 8885-KGB
; TELEPHONE: (914) 332-1700
; TELEFAX: (914) 332-1844
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2233 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Candida albicans
; US-08-145-705A-1

Alignment Scores:
Pred. No.: 0.187 Length: 2233
Score: 75.00 Matches: 25
Percent Similarity: 47.56% Conservative: 14
Best Local Similarity: 30.49% Mismatches: 43
Query Match: 9.65% Indels: 0
DB: 1 Gaps: 0

US-09-847-539A-6 (1-159) x US-08-145-705A-1 (1-2233)

QY 31 ArgAsnGluGluArgAlaIleAspGluLeuLysLysGlnAlaIleGluAspLysGluAla 50
Db 315 AAACGTGAAGAAGAGGCTAAGAGAAGGAAGAGAGGCCCAAGAGGAAGAGAGGCC 256
QY 51 ThrThrAlaIleGluAlaIleAspSerAspAlaLeuGluAlaLeuAlaAspGlnThrAsp 70
Db 255 AAGAAGAAGGAAGAGGCCCAAGAGAAGGCAGAGAGGCCCAAGAGAGAGAGGAG 196
QY 71 AlaLeuGlnSerGluGluAlaAlaValLysAlaAspAsnAlaIleSerAspAlaLeu 90
Db 195 GCCAAGAAGAGGCGAGAGAGGCCCAAGAGAAGGCAGAGAGGCCCAAGAGAGG 136
QY 91 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaGluValGlnSer 110
Db 135 GAGGCCAAGAAGAGGCGAGAGGCCCAAGAGAAGGCAGAGAGGCCCGTCCCAAAAAGTTT 76
QY 111 AspAsn 112
Db 75 GACAAC 70

RESULT 6
US-08-139-937-12
; Sequence 12, Application US/08139937
; Patent No. 5821070
; GENERAL INFORMATION:
; APPLICANT: LEE, WEN-HWA
; APPLICANT: SHAN, BEI
; TITLE OF INVENTION: CELLULAR GENES ENCODING
; TITLE OF INVENTION: RETINOBLASTOMA-ASSOCIATED PROTEINS
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CAMPBELL AND FLORES
; STREET: 4370 LA JOLLA VILLAGE DRIVE
; CITY: SAN DIEGO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:

; CLASSIFICATION: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/139,937
; FILING DATE: 20-OCT-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/979,156
; FILING DATE: 20-NOV-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: CAMPBELL, CATHRYN
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-CJ 9370
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-535-9001
; TELEFAX: 619-535-8949
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4868 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-139-937-12

Alignment Scores:
Pred. No.: 0.533 Length: 4868
Score: 75.00 Matches: 21
Percent Similarity: 49.00% Conservative: 28
Best Local Similarity: 21.00% Mismatches: 51
Query Match: 9.65% Indels: 0
DB: 1 Gaps: 0

US-09-847-539A-6 (1-159) x US-08-139-937-12 (1-4868)

QY 46 GluAspLysGluAlaThrThrAlaIleGluAlaIleSerAspAlaLeuGluAlaLeu 65
Db 2016 GAAAAGAGCAAGAGAAAGTACAGATGAACAGAAAATCAAGCACTGCCATCGAGATGCTT 2075
QY 66 AlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLysAlaAspAsnAla 85
Db 2076 CAACACAAATTAAGAGCTCAATGAGAGAGTGGCAGCCCTGCATATATGACCAAGAGGCC 2135
QY 86 AlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAla 105
Db 2136 TGTAAGCCCAAGAGCAGCAATCTTAGTAGTCAAGTAGAGTGTCTTGAAGTGGAGAGGCT 2195
QY 106 GluValValGlnSerAspAsnAlaIleSerAspAlaTrpGluLysAlaAlaThrProIle 125
Db 2196 CAGTTGCTACAGGCCCTTGATGAGGCCAAAATAATATATATTGTTTTCATCTTCAGTG 2255
QY 126 AlaLeuAspValLysLysThrLysAspThrLysProValLysLysGluGluArgGln 145
Db 2256 AATGGCCTCATTTCAAGAGTAGAAGTGGCAAGCAAGCAAACTGGAGAGAGGATGAAGAA 2315

RESULT 7
PCT-US93-11310-12
; Sequence 12, Application PC/TUS9311310
; GENERAL INFORMATION:
; APPLICANT: BOARD OF REGENTS OF THE UNIVERSITY OF TEXAS SYSTEM
; TITLE OF INVENTION: CELLULAR GENES ENCODING
; TITLE OF INVENTION: RETINOBLASTOMA-ASSOCIATED PROTEINS
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CAMPBELL AND FLORES
; STREET: 4370 LA JOLLA VILLAGE DRIVE
; CITY: SAN DIEGO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
```

```
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10136 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: HUMAN
US-08-353-700-2

Alignment Scores:
Pred. No.: 1.01 Length: 10136
Score: 76.00 Matches: 21
Percent Similarity: 49.00% Conservative: 28
Best Local Similarity: 21.00% Mismatches: 51
Query Match: 9.78% Indels: 0
DB: 1 Gaps: 0

US-09-847-539a-6 (1-159) x US-08-353-700-2 (1-10136)
QY 46 GluAspLysGluAlaThrThrAlaIleGluAlaSerSerAspAlaLeuGluAlaLeu 65
Db 7728 GAAAGAGCAAGAGAACTACAGATGAAAGAAATAATCAAGCACTGCCATGGAGATGCTT 7787
QY 66 AlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLysAlaAspAla 85
Db 7788 CAACACAAATTAAGAGCTCAATGAGAGAGTGCGCAGCCCTGCATATACCCAGAGGCC 7847
QY 86 AlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAla 105
Db 7848 TGTAAGGCCAAAGAGCAGAACTTAGTAGTCAAGTAGAGTGCTTCAACTTGAGAGGCT 7907
QY 106 GluValValGlnSerAspAlaAlaSerAspAlaTlpGluLysAlaAlaThrProIle 125
Db 7908 CAGTTGCTACAGGCTTATGAGGCCAAATAATATATATTTGCAATCTTCAGTG 7967
QY 126 AlaLeuAspValLysThrLysAspThrLysProValValLysGluGluArgGln 145
Db 7968 AAAGCCCTATTCAAGACTAGAGATGCGCAGCAAACTGGAGAGAGGATGAAGAA 8027

RESULT 4
PCT-US95-16216-2
; Sequence 2, Application PC/TUS9516216
; GENERAL INFORMATION:
; APPLICANT: Yen, Timothy J.
; TITLE OF INVENTION: Nucleic Acid Encoding a Transiently
; TITLE OF INVENTION: Expressed Kinetochores Protein, and Methods of Use
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dann, Dorfman, Herrell and Skillman
; STREET: 1601 Market Street Suite 720
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/16216
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/353,700
; FILING DATE: 09-DEC-1995
; ATTORNEY/AGENT INFORMATION:
```

```
; NAME: Reed, Janet E.
; REGISTRATION NUMBER: 36,252
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10136 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
PCT-US95-16216-2

Alignment Scores:
Pred. No.: 1.01 Length: 10136
Score: 76.00 Matches: 21
Percent Similarity: 49.00% Conservative: 28
Best Local Similarity: 21.00% Mismatches: 51
Query Match: 9.78% Indels: 0
DB: 5 Gaps: 0

US-09-847-539a-6 (1-159) x PCT-US95-16216-2 (1-10136)
QY 46 GluAspLysGluAlaThrThrAlaIleGluAlaSerSerAspAlaLeuGluAlaLeu 65
Db 7728 GAAAGAGCAAGAGAACTACAGATGAAAGAAATAATCAAGCACTGCCATGGAGATGCTT 7787
QY 66 AlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLysAlaAspAla 85
Db 7788 CAACACAAATTAAGAGCTCAATGAGAGAGTGCGCAGCCCTGCATATACCCAGAGGCC 7847
QY 86 AlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAla 105
Db 7848 TGTAAGGCCAAAGAGCAGAACTTAGTAGTCAAGTAGAGTGCTTCAACTTGAGAGGCT 7907
QY 106 GluValValGlnSerAspAlaAlaSerAspAlaTlpGluLysAlaAlaThrProIle 125
Db 7908 CAGTTGCTACAGGCTTATGAGGCCAAATAATATATATTTGCAATCTTCAGTG 7967
QY 126 AlaLeuAspValLysThrLysAspThrLysProValValLysGluGluArgGln 145
Db 7968 AAAGCCCTATTCAAGACTAGAGATGCGCAGCAAACTGGAGAGAGGATGAAGAA 8027

RESULT 5
US-08-145-705A-1/G
; Sequence 1, Application US/08145705A
; Patent No. 5489513
; GENERAL INFORMATION:
; APPLICANT: Springer, Wolfgang; Piempel, Manfred;
; APPLICANT: L. berding, Antonius
; TITLE OF INVENTION: SPECIFIC GENE PROBES AND
; TITLE OF INVENTION: PROCESSES FOR THE DIAGNOSTIC
; TITLE OF INVENTION: INVESTIGATION OF CANDIDA
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SPRUNG HORN KRAMER & WOODS
; STREET: 660 White Plains Road
; CITY: Tarrytown
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10591-5144
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.4 MB storage
; COMPUTER: NEC PowerMate 1 Plus
; OPERATING SYSTEM: DOS
; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/145,705A
; FILING DATE: October 28, 1993
```

```
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 61743/102
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1555 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 288..1526
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 288..1526
; US-08-669-408B-9
Alignment Scores:
Pred. No.: 0.000481 Length: 1555
Score: 91.00 Matches: 19
Percent Similarity: 62.50% Conservative: 1
Best Local Similarity: 59.38% Mismatches: 12
Query Match: 11.71% Indels: 0
DB: 3 Gaps: 0
US-09-847-539A-6 (1-159) x US-08-669-408B-9 (1-1555)
Qy 128 AspValLysLysThrLysAspThrLysProValValLysLysGluArgGlnAsnVal 147
Db 1356 GATCTAGAGAGAGATGCTAAAAAACCCAGAGCTAAGAAAGAAAGCTAAGAAAGCT 1415
Qy 148 AsnThrLeuProThrThrGlyGluGluSerAsnPro 159
Db 1416 GCAACTCTCTACACTGGTGAAGGAAGCAACCCA 1451
RESULT 2
; Sequence 1, Application US/08669408B
; Patent No. 6100055
; GENERAL INFORMATION:
; APPLICANT: GUSS, Bengt
; APPLICANT: JONSSON, Hans
; APPLICANT: LINDBERG, Martin
; APPLICANT: MUELLER, Hans-Peter
; APPLICANT: RANTAMAKI, Liisa K.
; TITLE OF INVENTION: METHOD AND MEANS FOR PRODUCING
; TITLE OF INVENTION: PLASMAPROTEINASE INHIBITOR-BINDING PROTEINS
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/669,408B
; FILING DATE: 03-JUL-1996
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/SE94/00826
; FILING DATE: 06-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: SE 9302855-3
```

```
; FILING DATE: 06-SEP-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 61743/102
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2526 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 506..2497
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 506..2497
; US-08-669-408B-1
Alignment Scores:
Pred. No.: 0.00719 Length: 2526
Score: 85.00 Matches: 18
Percent Similarity: 62.50% Conservative: 2
Best Local Similarity: 56.25% Mismatches: 12
Query Match: 10.94% Indels: 0
DB: 3 Gaps: 0
US-09-847-539A-6 (1-159) x US-08-669-408B-1 (1-2526)
Qy 128 AspValLysLysThrLysAspThrLysProValValLysLysGluArgGlnAsnVal 147
Db 2327 GACGCTAAAAAACCCAGAGCTAAGAAAGAAAGCTAAGAAAGAAAGCTAAGAAAGCT 2386
Qy 148 AsnThrLeuProThrThrGlyGluGluSerAsnPro 159
Db 2387 GCAACTCTCTACACTGGTGAAGGAAGCAACCCA 2422
RESULT 3
; Sequence 2, Application US/08353700
; Patent No. 5599919
; GENERAL INFORMATION:
; APPLICANT: YEN, TIMOTHY J.
; APPLICANT: RAYNER, JEROME B.
; TITLE OF INVENTION: NUCLEIC ACID ENCODING A
; TITLE OF INVENTION: TRANSIENTLY-EXPRESSED KINETOCHORE PROTEIN,
; TITLE OF INVENTION: AND METHODS OF USE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DANN, DORFMAN, HERRELL AND SKILLMAN
; STREET: 1601 MARKET STREET, SUITE 720
; CITY: PHILADELPHIA
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/353,700
; FILING DATE: 09-DEC-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: REED, JANET E.
; REGISTRATION NUMBER: 36,252
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
```

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 13, 2002, 02:17:36 ; Search time 51.8663 Seconds  
(without alignments)  
753.008 Million cell updates/sec

Title: US-09-847-539A-6  
Perfect score: 777  
Sequence: 1 VDSIEQRIIPNGTITNL.....KKERQNVNLTPTGERSNP 159

Scoring table: BLOSUM62  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 383533 seqs, 122816752 residues  
Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODE=frame+p2n.model -DEV=xlh  
-O=/cgn2\_1/USPTO\_spool/US09847539/runat\_10102002\_092549\_3456/app\_query.fasta\_1.526  
-DB=Issued\_Patents\_NA -QWTF=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=40  
-MODE=LOCAL -OUTFT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09847539-ecgn\_1\_13\_runat\_10102002\_092549\_3456 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_WMAP -LARGEOUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOPOP=6 -FGAPEXT=7  
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : Issued\_Patents\_NA:\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/PTUS\_COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	91	11.7	1555	3 US-08-669-408B-9	Sequence 9, Appli
2	85	10.9	2526	3 US-08-669-408B-1	Sequence 1, Appli
3	76	9.8	10136	1 US-08-353-700-2	Sequence 2, Appli
4	76	9.8	10136	5 PCT-US95-16216-2	Sequence 2, Appli
5	75	9.7	2233	1 US-08-145-705A-1	Sequence 1, Appli
6	75	9.7	4868	1 US-08-139-937-12	Sequence 12, Appli
7	75	9.7	4868	5 PCT-US93-11310-12	Sequence 12, Appli
8	75	9.7	8789	1 US-08-328-254-5	Sequence 5, Appli
9	71	9.1	605	4 US-08-936-165A-93	Sequence 93, Appli
10	68	8.8	5361	4 US-08-973-462-2	Sequence 2, Appli
11	68	8.8	6152	4 US-08-973-462-1	Sequence 1, Appli
12	67	8.6	1419	2 US-08-216-894-9	Sequence 9, Appli

13	67	8.6	1419	4 US-09-115-746-9	Sequence 9, Appli
14	66	8.5	2622	2 US-08-591-079-7	Sequence 7, Appli
15	66	8.5	3393	2 US-08-591-079-9	Sequence 9, Appli
16	65	8.4	3489	2 US-08-728-323A-1	Sequence 1, Appli
17	65	8.4	32207	2 US-08-770-379-20	Sequence 20, Appli
18	65	8.4	32207	4 US-08-757-669A-20	Sequence 20, Appli
19	65	8.4	32207	4 US-09-230-371A-20	Sequence 20, Appli
20	64	8.2	1695	2 US-08-216-894-1	Sequence 1, Appli
21	64	8.2	1695	4 US-09-115-746-1	Sequence 1, Appli
22	64	8.2	1932	2 US-08-216-894-7	Sequence 7, Appli
23	64	8.2	1932	4 US-09-115-746-7	Sequence 7, Appli
24	64	8.2	4316	1 US-08-317-450B-14	Sequence 14, Appli
25	64	8.2	4316	3 US-08-800-593-14	Sequence 14, Appli
26	64	8.2	5200	1 US-08-317-450B-12	Sequence 12, Appli
27	64	8.2	5200	3 US-08-800-593-12	Sequence 12, Appli
28	63	8.1	867	2 US-08-961-858-2	Sequence 2, Appli
29	63	8.1	867	3 US-08-961-858-4	Sequence 4, Appli
30	63	8.1	867	3 US-09-089-593-2	Sequence 2, Appli
31	63	8.1	867	3 US-09-089-593-3	Sequence 4, Appli
32	63	8.1	867	3 US-08-950-925-3	Sequence 3, Appli
33	62	8.0	397	3 US-09-253-691-3	Sequence 3, Appli
34	62	8.0	1890	4 US-09-346-408-7	Sequence 7, Appli
35	62	8.0	3252	2 US-08-770-301A-4	Sequence 4, Appli
36	62	8.0	4248	3 US-08-175-581-4	Sequence 1, Appli
37	62	8.0	4248	3 US-08-678-614-1	Sequence 1, Appli
38	62	8.0	4378	2 US-09-080-897-3	Sequence 3, Appli
39	62	8.0	4378	4 US-09-323-735-3	Sequence 3, Appli
40	62	8.0	4399	3 US-08-899-595-2	Sequence 2, Appli
41	62	8.0	6744	1 US-08-119-125A-2	Sequence 2, Appli
42	62	8.0	8878	1 US-08-759-444-2	Sequence 2, Appli
43	62	8.0	9880	3 US-08-680-897-1	Sequence 1, Appli
44	61	7.9	456	3 US-09-080-855-30	Sequence 30, Appli
45	61	7.9	2360	1 US-08-466-390-1	Sequence 1, Appli

## ALIGNMENTS

RESULT 1  
US-08-669-408B-9  
; Sequence 9, Application US/08669408B  
; Patent No. 6100055  
; GENERAL INFORMATION:  
; APPLICANT: GUSS, Bengt  
; APPLICANT: JONSSON, Hans  
; APPLICANT: LINDBERG, Martin  
; APPLICANT: MUELLER, Hans-Peter  
; APPLICANT: RANTAMAKI, Liisa K.  
; TITLE OF INVENTION: METHOD AND MEANS FOR PRODUCING  
; PLASMAPROTEINASE INHIBITOR-BINDING PROTEINS  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 3000 K Street, N.W., Suite 500  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20007-5109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION NUMBER: US/08/669.408B  
; FILING DATE: 03-JUL-1996  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: WO PCT/SE94/00826  
; FILING DATE: 06-SEP-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: SE 9302855-3  
; FILING DATE: 06-SEP-1993  
; ATTORNEY/AGENT INFORMATION:

```

|||||  |||  |||  |||  |||
Db  548 ACGGCTGCGCATCGTCAGCCAGTTTCGGCGCTCATCGGCATCATCGGCGCTGCTTCA 489
Qy  103 GluGluAlaGluValValGlnSerAsnAlaAlaSerAlaTrpGluLysAlaAla 122
Db  488 AAAGATGAGCGCACCGAGTCAGCAGCGGAGAGACAGCGCCACCGGATGCTCCGTG 429
Qy  123 ThrProIleAlaLeuAspVal 129
Db  428 GCCTCTGTGGCCTCCTCTGTG 408
```

Search completed: October 13, 2002, 02:25:02  
Job time : 241.572 secs

PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX  
PS Claim 1: SEQ ID NO 25641; 103pp; English.  
XX  
XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AA564197-AA594564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 3069 BP; 818 A; 832 C; 784 G; 635 T; 0 other;

Alignment Scores:  
Pred. No.: 0.212 Length: 3069  
Score: 81.00 Matches: 26  
Percent Similarity: 43.01% Conservative: 14  
Best Local Similarity: 27.96% Mismatches: 53  
Query Match: 10.42% Indels: 0  
DB: 23 Gaps: 0  
US-09-847-539A-6 (1-159) x AAS89837 (1-3069)  
QY 50 AlaThrThrAlaIleGluAlaIleSerSerAlaLeuGluAlaLeuAlaAspGlnThr 69  
DB 157 GCACGAGGTCGCTCCGCGAGCGAAGACATCCGAAAGCTACGCGAAGCGCGAACA 216  
QY 70 AspAlaLeuGlnSerGluAlaAlaValValValValValValValValValVal 89  
DB 217 AGCGCAGAAATCCCAAAAACGGCTCGCGCATCGTCAGCCAGTTTCGGCGGCTCATCGCA 276  
QY 90 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluAlaGluValValGln 109  
DB 277 TCATCGCGTCGCTCAAAAGATGAGGCGACAGACAGACAGCGTACGCGAAGCGAGC 336  
QY 110 SerAspAsnAlaAlaSerAspAlaTrpGluLysAlaAlaThrProIleAlaLeuAspVal 129  
DB 337 GCCACGACGGCATCCACGAAGGCGACAGAGGCTGCTGGCAGTCGCGGCGGCGCTCAG 396  
QY 130 LysLysThrLysAspThrLysProValValLysLysGlu 142  
DB 397 AGCAAAAGTACGCGGGAATCCGCGGCAACGCGCGCGAG 435

RESULT 40  
AAS74086/C  
ID AAS74086 standard; cDNA; 782 BP.  
XX  
AC AAS74086;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE DNA encoding novel human diagnostic protein #9890.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX Homo sapiens.  
OS WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
XX 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX (HYSE-) HYSEQ INC.  
XX  
PI Dmanac RT, Liu C, Tang YT;  
XX  
XX WPI; 2001-639362/73.  
DR P-PSDB; ABC09899.  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
PS Claim 1: SEQ ID NO 9890; 103pp; English.  
XX  
XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AA564197-AA594564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 782 BP; 135 A; 238 C; 246 G; 163 T; 0 other;

Alignment Scores:  
Pred. No.: 0.131 Length: 782  
Score: 77.00 Matches: 29  
Percent Similarity: 41.12% Conservative: 15  
Best Local Similarity: 27.10% Mismatches: 63  
Query Match: 9.91% Indels: 0  
DB: 23 Gaps: 0  
US-09-847-539A-6 (1-159) x AAS74086 (1-782)  
QY 23 AsnAlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaIleAspGluLeuLysLys 42  
DB 728 TCAGCCAGTCCACGCCCGCGGGAATATCGGACAGAAAGCGCCGATCGTCTTCAACA 669  
QY 43 GlnAlaIleGluAspLysGluAlaThrThrAlaIleGluAlaIleSerAspAlaLeu 62  
DB 668 GCCACAACGAAGGCTGCGGAAGCCTGCTGAAACAGCCAGCGAGCGTCTGCTTCC 609  
QY 63 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaValValValAla 82  
DB 608 GCAGCGAAGACATCCGAAACGAGCGGGAAGAGCGGGAACGAGCGGAGTCTCTCAAAA 549  
QY 83 AspAsnAlaIleSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSer 102

CC infections, abscesses and spontaneous abortion. This sequence represents  
CC DNA encoding the *S. dysgalactiae* Mig protein of the invention.

XX Sequence 2010 BP; 735 A; 393 C; 381 G; 501 T; 0 other;

#### Alignment Scores:

Pred. No.: 0.12 Length: 2010  
Score: 81.00 Matches: 17  
Percent Similarity: 62.50% Conservativity: 3  
Best Local Similarity: 53.12% Mismatches: 12  
Query Match: 10.42% Indels: 0  
DB: 24 Gaps: 0

US-09-847-539A-6 (1-159) x AAS18114 (1-2010)

QY 128 AspValLysThrLysAspThrLysProValValLysLysGluGluArgGlnAsnVal 147

DB 1837 GAAGCTAAACACGAGCTAAGAAAGAGAGCTAAGAAAGAGCTAAGAAAGCT 1896

QY 148 AsnThrLeuProThrThrGlyGluGluSerAsnPro 159

DB 1897 GCAACTCTTCTACAACTGGTGAAGGAAGCAACCCA 1932

#### RESULT 38

ID AAS77362 standard; cDNA; 3069 BP.

XX AAS77362;

DT 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #13166.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

DR WPI; 2001-639362/73.

DR P-PSDB; ABG13175.

XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity.

PS Claim 1; SEQ ID No 13166; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.

CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

SQ Sequence 3069 BP; 818 A; 832 C; 784 G; 635 T; 0 other;

#### Alignment Scores:

Pred. No.: 0.212 Length: 3069  
Score: 81.00 Matches: 26  
Percent Similarity: 43.01% Conservativity: 14  
Best Local Similarity: 27.96% Mismatches: 53  
Query Match: 10.42% Indels: 0  
DB: 23 Gaps: 0

US-09-847-539A-6 (1-159) x AAS77362 (1-3069)

QY 50 AlaThrThrAlaIleGluAlaAlaSerSerAspAlaLeuGluAlaLeuAlaAspGlnThr 69

DB 157 GCAGCGAGGTCGTCTCCGCGCAGGAGACATCCGAAACGTACGCGAAAGCGTCGGAACA 216

QY 70 AspAlaLeuGlnSerGluGluAlaAlaValValLysAlaAspAlaAlaSerAspAla 89

DB 217 ACCGCGAATCTCTCAAAACGGCTGCCGCTCAGGAGTTCGGCGGCGTCATCGGCA 276

QY 90 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaGluValValGln 109

DB 277 TCATCGGCGTCTGCTTCAAAAGATGAGCGCAGCAGACAGCGTCAGCAGCAAGAGCAGC 336

QY 110 SerAspAsnAlaAlaSerAspAlaTrpGluLysAlaAlaThrProIleAlaLeuAspVal 129

DB 337 GCCACGACGGCATCCACGAGGCGCAGAGGCTGCTGGCAGTTCGCGCGCGCAGCTCAG 396

QY 130 LysLysThrLysAspThrLysProValValLysLysGlu 142

DB 397 AGCAAAAGTACGGCGGAATCCCGCGCAACGCGCGCGAG 435

#### RESULT 39

AAS9837

ID AAS9837-standard; cDNA; 3069 BP.

XX AAS9837;

DT 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #25641.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

DR WPI; 2001-639362/73.

DR P-PSDB; ABG25650.

XX

```
OY 126 AlaLeuAspValLys 130
Db 501 GAAGCGGAGAAAG 515
RESULT 36
AAS82435/c
ID AAS82435 standard; cDNA; 2561 BP.
XX
AC AAS82435;
XX
DT 13-FEB-2002 (first entry)
DE
DE DNA encoding novel human diagnostic protein #18239.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.
KW
KW Homo sapiens.
OS
XX WO200175067-A2.
PN
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
PR
XX 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
PA
XX Drmanac RT, Liu C, Tang YT;
PI
XX WPI; 2001-639362/73.
DR
XX P-PSDB; ABG18248.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 1: SEQ ID No 18239; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2561 BP; 498 A; 708 C; 666 G; 699 T; 0 other;

Alignment Scores:
Pred. No.: 0.0847 Length: 2561
Score: 83.00 Matches: 26
Percent Similarity: 43.01% Conservative: 14
Best Local Similarity: 27.96% Mismatches: 53
Query Match: 10.68% Indels: 0
DB: 23 Caps: 0
```

US-09-847-539A-6 (1-159) x AAS82435 (1-2561)

OY 50 AlaThrThrAlaIleGluAlaAlaSerSerAspAlaLeuGluAlaLeuAlaAspGlnThr 69  
Db 1406 GCAGCGAGGTCTGCTCCGAGGAGAGACATCCGAAACGACGAAAGCGTCGGAACA 1347  
OY 70 AspAlaLeuGlnSerGluAlaAlaValValLysAlaAspAsnAlaAlaSerAspAla 89  
Db 1346 AGCGCAGATCCTCAAAAACGGCTGCCGCATCGTCAGCCAGTTCGGCGCGCTCATCGCA 1287  
OY 90 LeuGluAlaLeuAlaAspGlnThrAlaLeuGlnSerGluGluAlaGluValValGln 109  
Db 1286 TCATCGCGCTGCTGTTCAAAAGATGAGCGCAGCAGACGAGCGTCAGCAGCAGCAGC 1227  
OY 110 SerAspAsnAlaAlaSerAspAlaTrpGluLysAlaAlaThrProIleAlaLeuAspVal 129  
Db 1226 GCCACGACGGCATCCACGAGGCGAGAGGCTGCTGTCAGTGGCGGCGGACGCTCAG 1167  
OY 130 LysLysThrLysAspThrLysProValValLysLysGlu 142  
Db 1166 AGCAAAAGTACGGCGGAATCCGGCGCAACGGCGCCGAG 1128

RESULT 37  
AAS18114  
ID AAS18114 standard; DNA; 2010 BP.  
XX  
AC AAS18114;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Streptococcus dysgalactiae Mig DNA.  
XX  
XX Mig protein; Fc receptor; bacterial infection; Streptococcus agalactiae;  
KW streptococcal infection; mastitis; septicemia; meningitis; bacteraemia;  
KW impetigo; arthritis; urinary tract infection; abscess; antiinflammatory;  
KW spontaneous abortion; antibiotic; antiarthritic; gene therapy; ds.  
XX  
OS Streptococcus dysgalactiae.  
XX  
FH Key Location/Qualifiers  
FT CDS 1..2010  
FT /tag= a  
FT /product= "S. dysgalactiae Mig protein"  
XX  
XX WO200196380-A2.  
XX  
XX 20-DEC-2001.  
XX  
XX 11-JUN-2001; 2001WO-CA00837.  
XX  
XX 12-JUN-2000; 2000US-211016P.  
XX  
XX (UYSA-) UNIV SASKATCHEWAN.  
XX  
XX Potter AA, Bolton AJ, Song XM;  
XX  
XX WPI; 2002-106467/14.  
DR P-PSDB; AAU11350.  
XX  
XX Novel vaccine composition comprising Mig protein (Fc receptor) of  
PT Streptococcus dysgalactiae useful for treating or preventing  
PT streptococcal infection such as mastitis in vertebrates -  
XX  
XX Example 1; Fig 1; 60pp; English.  
XX  
XX The invention relates to a vaccine composition comprising a vehicle and  
CC an Fc receptor protein, Streptococcus dysgalactiae Mig protein. The  
CC sequence and composition are useful for treating or preventing a  
CC bacterial infection e.g. a streptococcal infection which causes mastitis  
CC in a vertebrate subject. The vaccine compositions can also be used to  
CC treat other streptococcal infections caused by S. agalactiae, such as  
CC septicemia, meningitis, bacteraemia, impetigo, arthritis, urinary tract



PR	30-JUN-1999;	99US-0141287.	PR	15-SEP-1999;	99US-0154018.
PR	01-JUL-1999;	99US-0141842.	PR	16-SEP-1999;	99US-0154039.
PR	01-JUL-1999;	99US-0142154.	PR	20-SEP-1999;	99US-0154779.
PR	02-JUL-1999;	99US-0142055.	PR	22-SEP-1999;	99US-0155139.
PR	06-JUL-1999;	99US-0142390.	PR	23-SEP-1999;	99US-0155486.
PR	08-JUL-1999;	99US-0142803.	PR	24-SEP-1999;	99US-0155659.
PR	09-JUL-1999;	99US-0142920.	PR	28-SEP-1999;	99US-0156458.
PR	12-JUL-1999;	99US-0142977.	PR	29-SEP-1999;	99US-0156596.
PR	13-JUL-1999;	99US-0143542.	PR	04-OCT-1999;	99US-0157117.
PR	15-JUL-1999;	99US-0143624.	PR	05-OCT-1999;	99US-0157753.
PR	15-JUL-1999;	99US-0144005.	PR	06-OCT-1999;	99US-0157865.
PR	16-JUL-1999;	99US-0144085.	PR	07-OCT-1999;	99US-0158029.
PR	16-JUL-1999;	99US-0144086.	PR	08-OCT-1999;	99US-0158232.
PR	19-JUL-1999;	99US-0144325.	PR	12-OCT-1999;	99US-0158369.
PR	19-JUL-1999;	99US-0144331.	PR	13-OCT-1999;	99US-0159293.
PR	19-JUL-1999;	99US-0144332.	PR	13-OCT-1999;	99US-0159294.
PR	19-JUL-1999;	99US-0144333.	PR	13-OCT-1999;	99US-0159295.
PR	19-JUL-1999;	99US-0144334.	PR	14-OCT-1999;	99US-0159329.
PR	19-JUL-1999;	99US-0144335.	PR	14-OCT-1999;	99US-0159330.
PR	20-JUL-1999;	99US-0144352.	PR	14-OCT-1999;	99US-0159331.
PR	20-JUL-1999;	99US-0144632.	PR	14-OCT-1999;	99US-0159637.
PR	20-JUL-1999;	99US-0144884.	PR	14-OCT-1999;	99US-0159638.
PR	21-JUL-1999;	99US-0144814.	PR	18-OCT-1999;	99US-0159584.
PR	21-JUL-1999;	99US-0145086.	PR	21-OCT-1999;	99US-0160741.
PR	21-JUL-1999;	99US-0145088.	PR	21-OCT-1999;	99US-0160767.
PR	22-JUL-1999;	99US-0145085.	PR	21-OCT-1999;	99US-0160768.
PR	22-JUL-1999;	99US-0145087.	PR	21-OCT-1999;	99US-0160770.
PR	22-JUL-1999;	99US-0145089.	PR	21-OCT-1999;	99US-0160814.
PR	22-JUL-1999;	99US-0145192.	PR	21-OCT-1999;	99US-0160815.
PR	23-JUL-1999;	99US-0145145.	PR	22-OCT-1999;	99US-0160980.
PR	23-JUL-1999;	99US-0145218.	PR	22-OCT-1999;	99US-0160981.
PR	23-JUL-1999;	99US-0145224.	PR	22-OCT-1999;	99US-0160989.
PR	26-JUL-1999;	99US-0145276.	PR	25-OCT-1999;	99US-0161404.
PR	27-JUL-1999;	99US-0145913.	PR	25-OCT-1999;	99US-0161405.
PR	27-JUL-1999;	99US-0145918.	PR	25-OCT-1999;	99US-0161406.
PR	27-JUL-1999;	99US-0145919.	PR	26-OCT-1999;	99US-0161359.
PR	28-JUL-1999;	99US-0145951.	PR	26-OCT-1999;	99US-0161360.
PR	02-AUG-1999;	99US-0146386.	PR	26-OCT-1999;	99US-0161361.
PR	02-AUG-1999;	99US-0146388.	PR	28-OCT-1999;	99US-0161920.
PR	02-AUG-1999;	99US-0146389.	PR	28-OCT-1999;	99US-0161992.
PR	03-AUG-1999;	99US-0147038.	PR	28-OCT-1999;	99US-0161993.
PR	04-AUG-1999;	99US-0147204.	PR	29-OCT-1999;	99US-0162142.
PR	04-AUG-1999;	99US-0147302.			
PR	05-AUG-1999;	99US-0147192.			
PR	05-AUG-1999;	99US-0147260.			
PR	06-AUG-1999;	99US-0147303.			
PR	06-AUG-1999;	99US-0147416.			
PR	09-AUG-1999;	99US-0147493.			
PR	09-AUG-1999;	99US-0147935.			
PR	10-AUG-1999;	99US-0148171.			
PR	11-AUG-1999;	99US-0148319.			
PR	12-AUG-1999;	99US-0148341.			
PR	13-AUG-1999;	99US-0148565.			
PR	13-AUG-1999;	99US-0148684.			
PR	16-AUG-1999;	99US-0149368.			
PR	17-AUG-1999;	99US-0149175.			
PR	18-AUG-1999;	99US-0149426.			
PR	20-AUG-1999;	99US-0149722.			
PR	20-AUG-1999;	99US-0149723.			
PR	20-AUG-199				

```

DT 11-OCT-1995 (first entry)
XX
DE Synthetic protein G gene, MD1.
XX
KW Synthetic; protein G; IgG; Fc receptors; heavy chain; leukaemia;
KW lymphoma; cancer; autoimmune disease; ds.
XX
OS Synthetic.
XX
PN WO9506125-A.
XX
PD 02-MAR-1995.
XX
PF 23-AUG-1994; 94WO-US09141.
XX
PR 23-AUG-1993; 93US-0110653.
XX
PA (IMMU-) APPLIED IMMUNE SCI INC.
XX
PI Lee YM, Okarma TB, Talib S;
XX
DR WPI: 1995-106854/14.
DR P-PSDB; AAR71127.
XX
PT New peptide(s) which bind the Fc region of an immunoglobulin -
PT comprising a nontotal portion of the amino acid sequence of
PT Protein A and/or Protein G
XX
PS Claim 20; Fig 4A; 6lpp; English.
XX
CC The sequence is that of a synthetic protein G gene, MD-1. The
CC proteins produced by MD-1 are capable of specifically binding
CC the constant region of the heavy chain of IgG in the same way as
CC neutral Fc receptors. They can be used to analyse the structure and
CC function of Fc receptors, as well as in antibody production, cell
CC culture, diagnosis and therapy. They can be used to treat disorders
CC such as leukaemia and lymphoma, cancer and immune disorders.
CC See also AAQ84882-7.
XX
SQ Sequence 627 BP; 205 A; 118 C; 145 G; 159 T; 0 other;

Alignment Scores:
Pred. No.: 0.0128 Length: 627
Score: 83.00 Matches: 17
Percent Similarity: 61.2% Conservative: 2
Best Local Similarity: 54.84% Mismatches: 12
Query Match: 10.68% Indels: 0
DB: 16 Gaps: 0

US-09-847-539A-6 (1-159) x AAQ84886 (1-627)
QY 128 AspValLysThrLysAspThrLysProValValLysGluGluArgGlnAsnVal 147
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 511 GATACATAAAAGAGATGCTAAAGAAACCGAAGCTGTAAGAAAGAGCTGCTAAAGAGCT 570
QY 148 AsnThrLeuProThrThrGlyGluGluSerAsn 158
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 571 GAAACTCTCCGACTACTGGTGAAGTTCTTAAC 603

RESULT 35
AAC38790
ID AAC38790 standard; DNA; 812 BP.
XX
AC AAC38790;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 22241.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX

```

```

OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 19-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.

```

```
Db 3586 AAGGAAGCCCGCAGCAGGAGGAGCCCTAGCAAGGTTGTGACTCTGCGGAACCCAAG 3745
Qy 104 GluAlaGluValValGlnSerAspAlaAlaSerAspAlaTrpGluLysAlaAlaThr 123
    |||||
Db 3746 GAGCAGAGAGCAATACGATGAGAGCGCCACGCCCGTCTTATGATGTTTCTACCGCC 3805
    |||||

Qy 124 ProfileAlaLeuaspValLys 131
    |||
Db 3806 CCGCATCCAATGATGTCAGCAAG 3829
    |||

RESULT 32
AAQ84884
ID AAQ84884 standard; cDNA; 141 BP.
AC
XX
XX
XX 11-OCT-1995 (first entry)
XX
XX Expression cassette for SG-3.
DE
XX Synthetic; protein G; IgG; Fc receptors; heavy chain; leukaemia;
KW lymphoma; cancer; autoimmune disease; ds.
XX
XX Synthetic.
OS
XX WO9506125-A.
XX
XX 02-MAR-1995.
XX
XX 23-AUG-1994; 94WO-US09141.
XX
XX 23-AUG-1993; 93US-0110653.
XX
XX (IMMU-) APPLIED IMMUNE SCI INC.
XX
XX Lee YM, Okarma TB, Talib S;
XX
XX WPI; 1995-106854/14.
XX
XX P-PSDB; AAR71125.
XX
XX New peptide(s) which bind the Fc region of an immunoglobulin -
PT comprising a nontotal portion of the amino acid sequence of
PT Protein A and/or Protein G
XX
XX Claim 20; Fig 2C; 61pp; English.
XX
XX The sequence is that of an expression vector constructed to produce
CC SG-3 (synthetic protein G). SG-3 is capable of specifically binding
CC the constant region of the heavy chain of IgG in the same way as
CC neutral Fc receptors. It can be used to analyse the structure and
CC function of Fc receptors, as well as in antibody production, cell
CC culture, diagnosis and therapy. They can be used to treat disorders
CC such as leukaemia and lymphoma, cancer and immune disorders.
CC See also AAQ84882-7.
XX
XX Sequence 141 BP; 59 A; 26 C; 29 G; 27 T; 0 other:
SQ
Alignment Scores:
Pred. No.: 0.00174 Length: 141
Score: 83.00 Matches: 17
Percent Similarity: 61.29% Conservative: 2
Best Local Similarity: 54.84% Mismatches: 12
Query Match: 10.68% Indels: 0
DB: 16 Gaps: 0

US-09-847-539A-6 (1-159) x AAQ84884 (1-141)

Qy 128 AspValLysLysThrLysAspThrLysProValLysLysGluArgGlnAsnVal 147
    |||
Db 48 GATACATAAAAGAAGATGCTAAAAAACCAGGAGCTAAAAAAGAAGATGCTAAAAAGCT 107
    |||

Qy 148 AsnThrLeuProThrThrGlyGluSerAsn 158
    |||
```

```
Db 108 GAAACTCTGCCGACTACTGTGAGGTTCTTAAC 140
RESULT 33
AAQ84887
ID AAQ84887 standard; cDNA; 603 BP.
XX
XX AAQ84887;
XX
XX 11-OCT-1995 (first entry)
XX
XX Chimeric synthetic protein G gene, MD2.
DE
XX Synthetic; protein G; IgG; Fc receptors; heavy chain; leukaemia;
KW lymphoma; cancer; autoimmune disease; ds.
XX
XX Synthetic.
OS
XX WO9506125-A.
XX
XX 02-MAR-1995.
XX
XX 23-AUG-1994; 94WO-US09141.
XX
XX 23-AUG-1993; 93US-0110653.
XX
XX (IMMU-) APPLIED IMMUNE SCI INC.
XX
XX Lee YM, Okarma TB, Talib S;
XX
XX WPI; 1995-106854/14.
XX
XX P-PSDB; AAR71128.
XX
XX New peptide(s) which bind the Fc region of an immunoglobulin -
PT comprising a nontotal portion of the amino acid sequence of
PT Protein A and/or Protein G
XX
XX Claim 20; Fig 4B; 61pp; English.
XX
XX The sequence is that of a synthetic protein G gene, MD-2. The
CC proteins produced by MD-2 are capable of specifically binding
CC the constant region of the heavy chain of IgG in the same way as
CC neutral Fc receptors. They can be used to analyse the structure and
CC function of Fc receptors, as well as in antibody production, cell
CC culture, diagnosis and therapy. They can be used to treat disorders
CC such as leukaemia and lymphoma, cancer and immune disorders.
CC See also AAQ84882-6.
XX
XX Sequence 603 BP; 218 A; 120 C; 126 G; 139 T; 0 other:
SQ
Alignment Scores:
Pred. No.: 0.0122 Length: 603
Score: 83.00 Matches: 17
Percent Similarity: 61.29% Conservative: 2
Best Local Similarity: 54.84% Mismatches: 12
Query Match: 10.68% Indels: 0
DB: 16 Gaps: 0

US-09-847-539A-6 (1-159) x AAQ84887 (1-603)

Qy 128 AspValLysLysThrLysAspThrLysProValLysLysGluArgGlnAsnVal 147
    |||
Db 487 GATACATAAAAGAAGATGCTAAAAAACCAGGAGCTAAAAAAGAAGATGCTAAAAAGCT 546
    |||

Qy 148 AsnThrLeuProThrThrGlyGluSerAsn 158
    |||
Db 547 GAAACTCTGCCGACTACTGTGAGGTTCTTAAC 579

RESULT 34
AAQ84886
ID AAQ84886 standard; cDNA; 627 BP.
XX
XX AAQ84886;
XX
```



CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. AAS84197-AAS94564 represent novel human  
 CC diagnostic coding sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 3411 BP; 921 A; 898 C; 918 G; 674 T; 0 other;

#### Alignment Scores:

Pred. No.:	0.0451	Length:	3411
Score:	86.00	Matches:	26
Percent Similarity:	43.30%	Conservative:	16
Best Local Similarity:	26.80%	Mismatches:	55
Query Match:	11.07%	Indels:	0
DB:	23	Gaps:	0

US-09-847-539A-6 (1-159) x AAS82434 (1-3411)

Qy 46 GluAspLysGluAlaThrThrAlaIleGluAlaSerSerAspAlaLeuGluAlaLeu 65  
 Db 2026 CAGTCCAGCGCAGCGAGTCTGCTTCCGCGAGAGACATCCGAAACGAGCGGAAA 2085  
 Qy 66 AlaAspGlnThrAspAlaLeuGlnSerGluGluAlaValLysAlaAspAla 85  
 Db 2086 CGCTCGGAACAGCGGAGATCTCTAAACGCGTCCGCGATCGTACGCCAGTCCGCGG 2145  
 Qy 86 AlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAla 105  
 Db 2146 CGGTCATCGGCATCATCGCGCTGCTTCAAGATGAGGCGACACAGCGTCAGCA 2205  
 Qy 106 GluValValGlnSerAspAlaAlaSerAspAlaTrpGluLysAlaAlaThrProIle 125  
 Db 2206 GCGAAGACGCGCGCACGCGCGATCCACGAGGCGACAGAGGCTGTCGCGAGTCCGACG 2265  
 Qy 126 AlaLeuAspValLysThrLysAspThrLysProValValLysLysGlu 142  
 Db 2266 CGCGAGCTCAGCAAGAAAGTACGGCGAATCCGCGGCAACGCGCGCGAG 2316

#### RESULT 29

AAQ89196  
 ID AAQ89196 standard; DNA; 2526 BP.

XX AAQ89196;

XX 22-SEP-1995 (first entry)

XX S. dysgalactiae mig gene.  
 KW MIG; fast alpha-2-macroglobulin binding protein; FAM;  
 KW plasma proteinase-inhibitor binding protein; ss.  
 XX Streptococcus dysgalactiae strain SCL.

Key	Location/Qualifiers
FT promoter	394..399
FT	/*tag= a
FT promoter	/note= "putative promoter region"
FT	416..421
FT	/*tag= b
FT promoter	/note= "putative promoter region"
FT	447..453
FT	/*tag= c
FT promoter	/note= "putative promoter region"
FT	467..472
FT	/*tag= d
FT promoter	/note= "putative promoter region"
FT	494..499

FT RBS

FT CDS /\*tag= e  
 FT 506..2500  
 FT /\*tag= f  
 FT 506..595  
 FT /\*tag= g  
 FT 596..2497  
 FT /\*tag= h

XX WO9507296-A.

XX 16-MAR-1995.

XX 06-SEP-1994; 94WO-SE00826.

XX 06-SEP-1993; 93SE-0002855.

XX (GUSS/) GUSS B.

XX (JONS/) JONSSON H.

XX (LIND/) LINDBERG M.

XX (MOEL/) MUELLER H.

XX (RANT/) RANTAMAKI L K.

XX Guss B, Jonsson H, Lindberg M, Mueller H, Rantamaki LK;  
 PI WPI; 1995-123382/16.

XX P-PSDB; AAR71929.

XX DNA encoding fast alpha 2-macroglobulin-binding proteins - used  
 PT to obtain prods for sepn., detection or quantification or for  
 PT binding inhibition

XX Disclosure; Page 31; 50pp; English.

XX A phage lambda GEM-11 library of S. dysgalactiae SCL DNA was analyzed  
 CC for fast-acting alpha 2-macroglobulin (FAM)- and IgG-binding  
 CC activity. One clone, which expressed both activities, was  
 CC analyzed to obtain DNA encoding the FAM-binding protein, MIG  
 CC (AAR71929). The mig gene is given in AAQ89196.

XX Sequence 2526 BP; 924 A; 449 C; 470 G; 683 T; 0 other;

#### Alignment Scores:

Pred. No.:	0.0423	Length:	2526
Score:	85.00	Matches:	18
Percent Similarity:	62.50%	Conservative:	2
Best Local Similarity:	56.25%	Mismatches:	12
Query Match:	10.94%	Indels:	0
DB:	16	Gaps:	0

US-09-847-539A-6 (1-159) x AAQ89196 (1-2526)

Qy 128 AspValLysLysThrLysAspThrLysProValValLysLysGluLysGlnAsnVal 147  
 Db 2327 GACGCTAAACCCAGAGCTAAGAAAGAAAGCTAAGAAAGAAAGCTAAGAAAGCT 2386

Qy 148 AsnThrLeuProThrThrGlyGluSerAsnPro 159  
 Db 2387 GCAACTCTCTCACTGCTGAGGAAAGCAACCCA 2422

#### RESULT 30

ABL03677

ID ABL03677 standard; cDNA; 2129 BP.

XX ABL03677;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 5513.  
 DE Drosophila: developmental biology; cell signalling; insecticide;  
 KW pharmaceutical; gene; ss.  
 XX Drosophila melanogaster.





```
PS Disclosure; Page 29-30; 38pp; English.
XX A phage lambda EMBL3 library of S. dysgalactiae 8215 DNA was
CC analyzed for alpha-2-macroglobulin-, BSA- and IgG-binding activity.
CC Clone lambda-SD1, which expressed all 3 activities, was analyzed to
CC obtain DNA encoding the SA-binding protein, MAG. The mag gene
CC (given in Q86080) encoded a 44 kDa MAG protein (R71670).
CC Recombinant, immobilized MAG was used for serum albumin affinity
CC purification, detection and assay.
XX
XX Sequence 1555 BP; 582 A; 242 C; 305 G; 426 T; 0 other;

Alignment Scores:
Pred. No.: 0.0029 Length: 1555
Score: 91.00 Matches: 19
Percent Similarity: 62.50% Conservative: 1
Best Local Similarity: 59.38% Mismatches: 12
Query Match: 11.71% Indels: 0
DB: 16 Gaps: 0

US-09-847-539A-6 (1-159) x AAQ86080 (1-1555)
QY 128 AspValLysThrLysAspThrLysProValLysLysGluGluArgGlnAsnVal 147
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1356 GATACTAAGAAAGAGATGCTAAACCAAGAGCTAAGAAAGAGCTAAGAAAGCT 1415

QY 148 AsnThrLeuProThrGlyGluGluSerAsnPro 159
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1416 GCAACTCTTCTCACTGCTGAAGGAGCAACCCA 1451

RESULT 24
AAQ89197
ID AAQ89197 standard; DNA; 1555 BP.
XX
XX AAQ89197;
XX
XX 22-SEP-1995 (first entry)
XX
XX S. dysgalactiae mag gene.
DE
DE Mag; fast alpha-2-macroglobulin binding protein; FAM;
KW plasma proteinase-inhibitor binding protein; ss.
XX
XX Streptococcus dysgalactiae strain 8215.
XX
XX Key Location/Qualifiers
FH promoter 198..203
FT /*tag= a
FT /*note= "putative promoter region"
FT promoter 230..235
FT /*tag= b
FT promoter 249..254
FT /*note= "putative promoter region"
FT promoter 264..269
FT /*tag= c
FT /*note= "putative promoter region"
FT promoter 276..281
FT /*tag= d
FT RBS 288..1527
FT /*tag= e
FT CDS 288..1527
FT /*tag= f
FT sig_peptide 288..389
FT /*tag= g
FT mat_peptide 390..1524
FT /*tag= h
XX
XX WO9507296-A.
XX
XX 16-MAR-1995.
XX
XX 06-SEP-1994; 94WO-SE00826.
XX
```

```
PR 06-SEP-1993; 93SE-0002855.
XX
XX (GUSS/) GUSS B.
PA (JONS/) JONSSON H.
PA (LIND/) LINDBERG M.
PA (MUEL/) MUELLER H.
PA (RANT/) RANTAMAKI L K.
XX
XX Guss B, Jonsson H, Lindberg M, Mueller H, Rantamaki LK;
XX
XX WPI; 1995-123382/16.
DR P-PSDB; AAR71928.
XX
XX DNA encoding fast alpha 2-macro:globulin-binding proteins - used
PT to obtain prods. for seph., detection or quantification or for
PT binding inhibition
XX
XX Disclosure; Fig.4; 50pp; English.
XX
XX A phage lambda EMBL library of S. dysgalactiae 8215 DNA was analyzed
CC for fast-acting alpha-2-macroglobulin (FAM)-, BSA- and IgG-binding
CC activity. Clone lambda-SD1, which expressed all 3 activities, was
CC analyzed to obtain DNA encoding the FAM-binding protein, MAG. The
CC mag gene (given in Q89197) encodes a 44 kDa protein (R71929).
XX
XX Sequence 1555 BP; 583 A; 242 C; 304 G; 426 T; 0 other;

Alignment Scores:
Pred. No.: 0.0029 Length: 1555
Score: 91.00 Matches: 19
Percent Similarity: 62.50% Conservative: 1
Best Local Similarity: 59.38% Mismatches: 12
Query Match: 11.71% Indels: 0
DB: 16 Gaps: 0

US-09-847-539A-6 (1-159) x AAQ89197 (1-1555)
QY 128 AspValLysThrLysAspThrLysProValLysLysGluGluArgGlnAsnVal 147
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1356 GATACTAAGAAAGAGATGCTAAACCAAGAGCTAAGAAAGAGCTAAGAAAGCT 1415

QY 148 AsnThrLeuProThrGlyGluGluSerAsnPro 159
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1416 GCAACTCTTCTCACTGCTGAAGGAGCAACCCA 1451

RESULT 25
AAQ89197
ID AAQ89197 standard; DNA; 1469 BP.
XX
XX AAQ89197;
XX
XX 21-MAY-1991 (first entry)
XX
XX Sequence encoding polypeptide possessing IgG-binding activity of
DE protein G from Streptococcus G148.
XX
XX Antibody-binding; IgG; IgA; immunoglobulin; ss.
XX
XX Streptococcus G148.
XX
XX Key Location/Qualifiers
FH CDS 1..1443
FT /*tag= a
FT misc_feature 9..80
FT /*tag= b
FT /*label= A1
FT misc_feature 81..233
FT /*tag= c
FT /*label= B1
FT misc_feature 234..305
FT /*tag= d
FT /*label= A2
FT misc_feature 306..458
```



```
RESULT 22
ID AAS80225 standard; cDNA; 1383 BP.
XX
AC AAS80225;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #16029.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
XX P-PSDB; ABG16038.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity
XX
XX Claim 1; SEQ ID NO 16029; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. AAS84197-AAS94564 represent novel human
XX diagnostic coding sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 1383 BP; 331 A; 349 C; 439 G; 264 T; 0 other;

Alignment Scores:
Pred. No.: 0.00126 Length: 1383
Score: 93.00 Matches: 24
Percent Similarity: 54.41% Conservative: 13
Best Local Similarity: 35.29% Mismatches: 31
Query Match: 11.97% Indels: 0
DB: 23 Gaps: 0

US-09-847-539A-6 (1-159) x AAS80225 (1-1383)

QY 50 AlaThrThrAlaIleGluAlaAlaSerSerAspAlaLeuGluAlaLeuAlaAspGlnThr 69
```

```
Db 1162 GCAGAGTCTCTCAAAAACGCGCGCCACCAAGTCCGCGTCCGCGGAAACGTCAGAAACG 1221
QY 70 AspAlaLeuGlnSerGluGluAlaAlaValVallyAlaAlaAspAsnAlaAlaSerAspAla 89
Db 1222 AATGCTGCACCGTCACAACAATCAGCGCCACGCTCTGCTCCACCGCGGCAGAAAGCG 1281
QY 90 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaGluValValGln 109
Db 1282 TCAGAGGCCGCCACTTCAGCACGAGATGCGTGGCTCAAAAGAGCGCAGCAAAATCATCA 1341
QY 110 SerAspAsnAlaAlaSerAspAla 117
Db 1342 GAAACGAAACGATCATCAAGTGCC 1365

RESULT 23
AAO86080
ID AAO86080 standard; DNA; 1555 BP.
XX
AC AAO86080;
XX
DT 21-SEP-1995 (first entry)
XX
DE S. dysgalactiae mag gene.
XX
KW Mag gene; serum albumin binding protein; protein stabilization;
KW vaccine; ss.
XX
OS Streptococcus dysgalactiae strain 8215.
XX
XX Key Location/Qualifiers
FH promoter 198..203
FT /tag= a
FT /note= "putative promoter sequence"
FT promoter 230..235
FT /tag= b
FT /note= "putative promoter sequence"
FT promoter 249..254
FT /tag= c
FT /note= "putative promoter sequence"
FT promoter 264..269
FT /tag= d
FT /note= "putative promoter sequence"
FT RBS 276..281
FT CDS 288..1527
FT /tag= e
FT /tag= f
FT sig_peptide 288..389
FT /tag= g
FT mat_peptide 390..1524
FT /tag= h
XX
XX WO9507300-A.
XX
XX 16-MAR-1995.
XX
XX 06-SEP-1994; 94WO-SR00825.
XX
XX 06-SEP-1993; 93SE-0002856.
XX
XX (GUSS/) GUSS B.
XX (JONS/) JONSSON H.
XX (LIND/) LINDBERG M.
XX
XX Guss B, Jonsson H, Lindberg M;
XX
XX WPI; 1995-123386/16.
XX P-PSDB; AAR71670.
XX
XX DNA encoding a serum albumin binding protein - used to obtain
XX prods. for sepsis., detection, quantification, protein
XX stabilisation or vaccine development
XX
```

```
FT sig_peptide 1..90
FT /*tag= a
FT /note= "encodes 30 amino acid secretion sequence"
FT 91..798
FT mat_peptide /*tag= b
FT /product= Protein_G_variant
XX US5312901-A.
XX PD 17-MAY-1994.
XX PF 14-FEB-1986; 86US-0829354.
XX PR 14-FEB-1986; 86US-0829354.
XX PR 23-APR-1986; 86US-0854887.
XX PR 19-JUN-1987; 87US-0063959.
XX PR 20-JUN-1988; 88US-0209236.
XX PR 19-JUN-1990; 90US-0540169.
XX PR 21-APR-1992; 92US-0871539.
XX (PHAA ) PHARMACIA LKB BIOTECHNOLOGY AB.
XX PI Fahnestock SR;
XX WPI; 1994-159179/19.
DR P-PSDB; AAR53294.
XX New recombinant streptococcal protein G variants - useful for
PT antibody detection and purification and for therapy
XX Disclosure; Columns 9-12 ; 48pp; English.
XX Protein G isolated from Streptococcus has IgG-binding activity which
CC has been localised to the B repeating structure (see AAR53290).
CC The nucleotide sequence AAQ64648 codes for a claimed Streptococcal
CC Protein G variant comprising the B domain active site and retaining
CC IgG-binding activity.
XX SQ Sequence 798 BP; 277 A; 161 C; 177 G; 183 T; 0 other;

Alignment Scores:
Pred. No.: 1.04e-05 Length: 798
Score: 105.00 Matches: 23
Percent Similarity: 72.73% Conservative: 1
Best Local Similarity: 69.70% Mismatches: 9
Query Match: 13.51% Indels: 0
DB: 15 Gaps: 0

US-09-847-539A-6 (1-159) x AAQ64648 (1-798)
QY 4 ProIleGluInProArgIleIleProAsnGlyGlyThrLeuThrAsnLeuGlyAsn 23
Db 100 CCAATCGAAGATACCCCAATTATTTCGTATGTTGTTGAATTAACATACTTCTGGGAAT 159
QY 24 AlaProGluLysLeuAlaAlaArgAsnGluGluArgAla 36
Db 160 TCAGAGACACACTGCCTTTCGTTAATGAAGAGCTGCT 198

RESULT 21
ID AAS84496
XX AAS84496 standard; cDNA; 1380 BP.
AC AAS84496;
XX 13-FEB-2002 (first entry)
XX DNA encoding novel human diagnostic protein #20300.
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX Homo sapiens.
OS
```

```
PN W0200175067-A2.
XX 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US08631.
XX PR 31-MAR-2000; 2000US-0540217.
XX PR 23-AUG-2000; 2000US-0649167.
XX (HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
DR P-PSDB; ABG20309.
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX Claim 1; SEQ ID No 20300; 103pp; English.
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 1380 BP; 331 A; 348 C; 438 G; 263 T; 0 other;

Alignment Scores:
Pred. No.: 0.00126 Length: 1380
Score: 93.00 Matches: 24
Percent Similarity: 54.41% Conservative: 13
Best Local Similarity: 35.29% Mismatches: 31
Query Match: 11.97% Indels: 0
DB: 23 Gaps: 0

US-09-847-539A-6 (1-159) x AAS84496 (1-1380)
QY 50 AlaThrThrAlaIleGluAlaAlaSerSerAspAlaLeuGluAlaLeuAlaAspGlnThr 69
Db 1159 GCAGAGTCTCTCAAAAACGGCGCCACCAGTGGCGTGGCGCGCAAAACGTGAGAAAGG 1218
QY 70 AspAlaLeuGlnSerGluAlaAlaValValLysAlaAspAsnAlaAlaSerAspAla 89
Db 1219 AATGCTGCAGCGTCACAACATCAGCGCCAGCTGCTCCAGCGCGGCCACGAAGG 1278
QY 90 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaValValGln 109
Db 1279 TCAGAGCGCGGCACCTTCAGCAGCATGCGGTGGCTCAAAAGAGGCGCAAAATCATCA 1338
QY 110 SerAspAsnAlaAlaSerAspAla 117
Db 1339 GAAACGACGCATCATCAAGTGCC 1362
```

Db 160 TCAGAGACAACACTGGCTTTGCGTAATGAAGAGAGTGCT 198  
::: ||||||||||||||||||| |||

## RESULT 18

AAQ06009

ID AAQ06009 standard; DNA; 798 BP.

AC AAQ06009;

XX 17-JAN-1991 (first entry)

XX Sequence encoding Protein G variant.

XX Immunoglobulin; ds.

XX Streptococcus sp. Lancefield Group G strain.

XX US4956296-A.

XX 11-SEP-1990.

XX 20-JUN-1988; 88US-0209236.

XX 20-JUN-1988; 88US-0209236.

XX 14-FEB-1986; 86US-0829354.

XX 23-APR-1986; 86US-0854887.

XX 17-FEB-1987; 87WO-US00329.

XX 19-JUN-1987; 87US-0063959.

XX (GENE-) GENEX CORP.

XX Fahnestock SR;

XX WPI; 1990-297491/39.

XX P-PSDB; AAR07004.

XX Recombinant Protein G variants - obt'd. using a cloned gene

XX encoding Protein G from Streptococcus sp., used for binding

XX immunoglobulin.

XX Claim 4; Column 47; 48pp; English.

XX Sequence may be incorporated into a non-pathogenic host eg. E.coli,

XX where they may be expressed at high levels. The proteins have a

XX higher binding efficiency and capacity for immunoglobulin, and may

XX be used for purifying, detecting and isolating antibodies.

XX Sequence 798 BP; 277 A; 161 C; 177 G; 183 T; 0 other;

SQ Alignment Scores:

Pred. No.: 1.04e-05 Length: 798

Score: 105.00 Matches: 23

Percent Similarity: 72.73% Conservative: 1

Best Local Similarity: 69.70% Mismatches: 9

Query Match: 13.51% Indels: 0

DB: 11 Gaps: 0

US-09-847-539A-6 (1-159) x AAQ06009 (1-798)

QY 4 ProtleGlulnProArGIllelleProAsnGlyGlyThrLeuThrAsnLeuGlyAsn 23

Db 100 CCAATCGAAGATACCCCAATTTTCGTAATGGTGAATTAACATAATCTCTGGGAAT 159

QY 24 AlaProGlulysLeuAlaLeuArgAsnGluGluArgAla 36

Db 160 TCAGAGACAACACTGGCTTTGCGTAATGAAGAGAGTGCT 198

::: ||||||||||||||||||| |||

RESULT 19

AAQ10007

ID AAQ10007 standard; DNA; 798 BP.

XX AAQ10007;

XX

XX

DT 13-MAR-1991 (first entry)

XX Type 4 GX7809 protein G variant.

XX Immunoglobulins; Ig; ds.

XX Streptococcus sp GX7809.

XX US4977247-A.

XX 11-DEC-1990.

XX 19-MAY-1989; 89US-0354264.

XX 19-MAY-1989; 89US-0354264.

XX 14-FEB-1986; 86US-0829354.

XX 23-APR-1986; 86US-0854887.

XX 17-FEB-1987; 87WO-US00329.

XX 19-JUN-1987; 87US-0063959.

XX 20-JUN-1988; 88US-0209236.

XX (GENE-) GENEX CORP.

XX Fahnestock SR, Lee T, Wroble MH;

XX WPI; 1991-006758/01.

XX P-PSDB; AAR10011.

XX Immobilised protein G variants - used for detection, isolation

XX and purificn. immunoglobulin(s) and immunoglobulin fragments

XX Disclosure; Column 11; 52pp; English.

XX Protein G variant product carries active regions B1 and B2, it may

XX be immobilised and exhibits different binding profiles. The bound

XX protein is useful in purification and detection of Igs and fragments.

XX Sequence 798 BP; 277 A; 161 C; 177 G; 183 T; 0 other;

SQ Alignment Scores:

Pred. No.: 1.04e-05 Length: 798

Score: 105.00 Matches: 23

Percent Similarity: 72.73% Conservative: 1

Best Local Similarity: 69.70% Mismatches: 9

Query Match: 13.51% Indels: 0

DB: 12 Gaps: 0

US-09-847-539A-6 (1-159) x AAQ10007 (1-798)

QY 4 ProtleGlulnProArGIllelleProAsnGlyGlyThrLeuThrAsnLeuGlyAsn 23

Db 100 CCAATCGAAGATACCCCAATTTTCGTAATGGTGAATTAACATAATCTCTGGGAAT 159

QY 24 AlaProGlulysLeuAlaLeuArgAsnGluGluArgAla 36

Db 160 TCAGAGACAACACTGGCTTTGCGTAATGAAGAGAGTGCT 198

::: ||||||||||||||||||| |||

RESULT 20

AAQ64648

ID AAQ64648 standard; DNA; 798 BP.

XX AAQ64648;

XX

XX 06-JAN-1995 (first entry)

XX IgG-binding Streptococcus Protein G variant coding sequence.

XX Streptococcus Protein G; variant; IgG binding activity;

XX immunoglobulin; Lancefield Group G; bacterial Fc receptor; ss.

XX Streptococcus sp.

XX Key

XX Location/Qualifiers

KW liposomal delivery complex; connecting moiety; antibody; Fc region;  
 KW diagnosis; therapeutic agent; antibiotic; antidepressant; antiviral;  
 KW antitumorigenic; cytokine; hormone; imaging agent; neurotransmitter;  
 KW stimulant; cytotoxic agent; malignant cell; ss.

XX Streptococcus sp.

XX Key Location/Qualifiers

FT CDS 292..603

FT FT /\*tag= a

FT FT 705..1262

FT FT /\*tag= b

FT FT /product= "protein G"

FT FT /transl\_except= (pos:705..707,aa:Met)

XX WO9959643-A2.

XX 25-NOV-1999.

XX 19-MAY-1999; 99WO-US11177.

XX 20-MAY-1998; 98US-0086347.

XX (SDGS-) SDG INC.

XX Lau JR;

XX WPI: 2000-062383/05.

XX P-PSDB; AAY57610, AAY57611.

XX New liposome constructs comprising a liposome connected to an antibody,  
 used, e.g. for delivery of cytotoxic agents to malignant cells -

XX Disclosure: Fig 1: 32pp; English.

XX The present invention describes a construct for connecting an antibody  
 or antibody fragment to a liposome. The construct comprises protein G'  
 and a linking moiety for connecting the protein G' to the liposome.  
 XX Also described are: (1) a liposomal delivery vehicle comprising: (a) a  
 liposome; and (b) a connecting moiety connected to the liposome, which  
 specifically binds the Fc region of an antibody, for connecting the  
 antibody to the liposome; and (2) forming a liposomal carrier comprising:  
 XX (a) providing liposomal components having binding moieties, for forming  
 a liposome; (b) providing a construct comprising a linking moiety and a  
 connecting moiety bound together; and (c) combining the liposomal  
 components and the construct and sonicating the combination so that the  
 binding moieties are exposed on a surface of the liposome, for binding to  
 the linking moiety, and to facilitate the binding between them. The  
 products can be used for the delivery of diagnostic or therapeutic  
 agents. The liposomes may contain or may be associated with a diagnostic  
 or therapeutic agent, e.g. antibiotics, antidepressants,  
 XX antitumorigenics, antivirals, cytokines, hormones, imaging agents,  
 neurotransmitters, or stimulants. They can be used particularly for the  
 delivery of cytotoxic agents to malignant cells. The protein G'  
 connecting moiety provides a liposomal delivery complex having improved  
 targeting efficiency. As a result of the binding between protein G' and  
 the Fc region of antibodies, protein G' shields the Fc regions of the  
 attached antibodies from non-specific binding to cell-surfaces, other  
 proteins, and anatomical structures. The present sequence encodes  
 XX an unidentified protein and protein G'.

XX Sequence 1576 BP; 572 A; 270 C; 330 G; 404 T; 0 other;

XX Alignment Scores:

Pred. No.:	9.57e-11	Length:	1576
Score:	142.00	Matches:	37
Percent Similarity:	60.61%	Conservative:	3
Best Local Similarity:	56.06%	Mismatches:	26
Query Match:	18.28%	Indels:	1
DB:	21	Gaps:	0

US-09-847-539A-6 (1-159) x AAZ47931 (1-1576)

QY 1 ValAspSerProIleGluInProArgIleIleProAsnGlyGlyThrLeuThrAsnLeu 20  
 |||||  
 Db 391 GTTGACTCAACCAATCGAAGATACCCCAATATTTCGTAATGGTGGTAATTAATCTT 450  
 |||||  
 QY 21 LeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaIleAspGluLeu 40  
 |||||  
 Db 451 CTGGGAATTCAGAGACAACACTGGCTTTCGCTAATGAAGAGAGTCTACAGCTGATTG 510  
 |||||  
 QY 41 LysLysGlnAlaIleGluAspLysGluAlaThrAlaIleGluAlaLaserSerAsp 60  
 |||||  
 Db 511 ACAGCAGCAGCGGTAGCCGATACTGTGGCAGCAGCGGACGCTGAAATGCTGGGCAGCA- 569  
 |||||

QY 61 AlaLeuGluAlaLeuAla 66

Db 570 GCTTGGGAAGCAGCGGCA 587

RESULT 17

AAAN91099

ID AAAN91099 standard; DNA; 798 BP.

XX AC AAAN91099;

XX 04-JUL-1990 (first entry)

XX Protein G gene variant.

XX Protein G; immunoglobulin; Fc receptor; ds.

XX Streptococcus sp.

XX WO8810306-A.

XX 29-DEC-1988.

XX 20-JUN-1988; 88WO-US02084.

XX 19-JUN-1987; 87US-0063959.

XX (GENE-) GENEX CORP.

XX Fahnestock SR;

XX WPI: 1989-023848/03.

XX P-PSDB; AAP94785.

XX Cloned protein G variant genes -

XX expressing proteins having immunoglobulin-binding properties of

XX protein G and derived from Streptococcus sp.

XX Claim 4; Page 83; 116pp; English.

XX Gene encodes protein G variant of non-pathogenic streptococcus allowing

XX isolation of the protein and variants, useful as bacterial Fc receptors

XX eg in purification and detection of Abs., screening of hybridoma clones

XX and treatment of disease.

XX Sequence 798 BP; 277 A; 161 C; 177 G; 183 T; 0 other;

XX Alignment Scores:

Pred. No.:	1.04e-05	Length:	798
Score:	105.00	Matches:	23
Percent Similarity:	72.73%	Conservative:	1
Best Local Similarity:	69.70%	Mismatches:	9
Query Match:	13.51%	Indels:	0
DB:	10	Gaps:	0

US-09-847-539A-6 (1-159) x AAAN91099 (1-798)

QY 4 ProfileGluInProArgIleIleProAsnGlyGlyThrLeuThrAsnLeuGlyAsn 23

Db 100 CCAATCGAAGATACCCCAATATTTCGTAATGGTGGTAATTAATCTTCTGGGAAT 159

QY 24 AlaProGluLysLeuAlaLeuArgAsnGluGluArgAla 36

```

XX (GENE-) GENEX CORP.
XX
XX Fahnstock SR, Lee T, Wroble MH;
XX
XX WPI; 1991-006758/01.
XX
XX P-PSDB; AAR10005.
XX
XX Immobilised protein G variants - used for detection, isolation
XX and purificn. immunoglobulin(s) and immunoglobulin fragments
XX
XX Disclosure; Fig 9; 52pp; English.
XX
XX Protein G gene product may be modified allowing the variant to
XX be immobilised and exhibit different binding profiles. The bound
XX protein is useful in purification and detection of Igs and fragments.
XX
XX Sequence 2383 BP; 867 A; 409 C; 481 G; 626 T; 0 other;

Alignment Scores:
Pred. No.:      2,05e-12      Length:      2383
Score:          155.00      Matches:      37
Percent Similarity: 62.12%      Conservative: 4
Best Local Similarity: 56.08%      Mismatches: 25
Query Match:     19.95%      Indels:      0
DB:              12          Gaps:          0

US-09-847-539A-6 (1-159) x AAQ10002 (1-2383)
QY 1 ValAspSerProIleGluGlnProArgIleIleProAsnGlyGlyThrLeuThrAsnLeu 20
Db 676 GTTGACTCACCACAGATACCCCAATATTGTAATGGTGGTAATACTAATCTT 735
QY 21 LeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaIleAspGluLeu 40
Db 736 CTGGGAATTCAGAGACAACACTGGCTTTGCGTAATGAAGAGAGTCTACAGCTGATTG 795
QY 41 LysLysGlnAlaIleGluAspLysGluAlaThrAlaIleGluAlaAlaSerSerAsp 60
Db 796 ACAGCAGCAGCGGTAGCCGATCTGTGGCAGCAGCGGAGCTGMAAATGCTGGGGCAGCA 855
QY 61 AlaLeuGluAlaLeuAla 66
Db 856 GCTTGGGAAGCAGCGGCA 873

RESULT 15
AAQ75036
ID AAQ75036 standard; DNA; 2383 BP.
XX
XX AC AAQ75036;
XX
XX 10-JAN-1995 (first entry)
XX
XX Streptococcus Protein G gene derived from strain GX7805.
XX
XX Streptococcus Protein G; variant; IgG binding activity;
XX immunoglobulin; Lancefield Group G; bacterial Fc receptor; ss.
XX
XX Streptococcus sp. GX7805.
XX
XX Key Location/Qualifiers
XX -35_signal 469
XX /*tag= a
XX -10_signal 486..491
XX /*tag= b
XX RBS 558..564
XX /*tag= c
XX CDS 577..2358
XX /*tag= d
XX /product= Protein_G
XX /transl_except= pos:1381..1383, aa:Gln
XX /transl_except= pos:1702..1704, aa:Leu
XX /transl_except= pos:1972..1974, aa:Gly

```

```

FT XX /transl_except= pos:2350..2352, aa:Gly
XX
XX US5312901-A.
XX
XX 17-MAY-1994.
XX
XX 14-FEB-1986; 86US-0829354.
XX
XX 14-FEB-1986; 86US-0829354.
XX
XX 23-APR-1986; 86US-0854887.
XX
XX 19-JUN-1987; 87US-0063959.
XX
XX 20-JUN-1988; 88US-0202336.
XX
XX 19-JUN-1990; 90US-0540169.
XX
XX 21-APR-1992; 92US-0871539.
XX
XX (PHAA ) PHARMACIA LKB BIOTECHNOLOGY AB.
XX
XX Fahnstock SR;
XX
XX WPI; 1994-159179/19.
XX
XX P-PSDB; AAR62944.
XX
XX New recombinant streptococcal protein G variants - useful for
XX antibody detection and purification and for therapy
XX
XX Example 5; Fig 9; 48pp; English.
XX
XX A 2.4kb HindIII fragment containing the entire coding sequence for
XX Protein G was isolated from Streptococcus GX7805 using the 1.9kb
XX Protein G coding sequence from Streptococcus GX7805. The protein G
XX has IgG-binding activity which has been localised to the B repeating
XX structure. Streptococcal Protein G variants comprising the B domains
XX are claimed.
XX
XX Sequence 2383 BP; 867 A; 409 C; 482 G; 625 T; 0 other;

Alignment Scores:
Pred. No.:      2,05e-12      Length:      2383
Score:          155.00      Matches:      37
Percent Similarity: 62.12%      Conservative: 4
Best Local Similarity: 56.08%      Mismatches: 25
Query Match:     19.95%      Indels:      0
DB:              15          Gaps:          0

US-09-847-539A-6 (1-159) x AAQ75036 (1-2383)
QY 1 ValAspSerProIleGluGlnProArgIleIleProAsnGlyGlyThrLeuThrAsnLeu 20
Db 676 GTTGACTCACCACAGATACCCCAATATTGTAATGGTGGTAATACTAATCTT 735
QY 21 LeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaIleAspGluLeu 40
Db 736 CTGGGAATTCAGAGACAACACTGGCTTTGCGTAATGAAGAGAGTCTACAGCTGATTG 795
QY 41 LysLysGlnAlaIleGluAspLysGluAlaThrAlaIleGluAlaAlaSerSerAsp 60
Db 796 ACAGCAGCAGCGGTAGCCGATCTGTGGCAGCAGCGGAGCTGMAAATGCTGGGGCAGCA 855
QY 61 AlaLeuGluAlaLeuAla 66
Db 856 GCTTGGGAAGCAGCGGCA 873

RESULT 16
AAZ47931
ID AAZ47931 standard; cDNA; 1576 BP.
XX
XX AC AAZ47931;
XX
XX 10-MAR-2000 (first entry)
XX
XX Streptococcus strain G 148 protein encoding cDNA.
XX
XX Streptococcus strain G 148; protein G'; protein G primer; liposome;

```

PT antibody detection and purification and for therapy

XX Example 2; Fig 3; 48pp; English.

XX A 1.9kb HindIII fragment containing the entire coding sequence for  
CC Protein G was isolated from Streptococcus GX7809. The protein G has  
CC IgG-binding activity which has been localised to the B repeating  
CC structure. Streptococcal Protein G variants comprising the B domains  
CC are claimed.

XX Sequence 1950 BP; 706 A; 323 C; 398 G; 523 T; 0 other;

Alignment Scores:  
Pred. No.: 1-57e-12 Length: 1950  
Score: 155.00 Matches: 37  
Percent Similarity: 62.12% Conservative: 4  
Best Local Similarity: 56.06% Mismatches: 25  
Query Match: 19.95% Indels: 0  
DB: 15 Gaps: 0

US-09-847-539A-6 (1-159) x AAQ064644 (1-1950)

QY 1 ValAspSerProIleGluGlnProArgIleIleProAsnGlyGlyThrLeuThrAsnLeu 20

DB 677 GTTGATTCCACCAATCGAAGATACCCCAATTATTGCGTAATGGTGGTGAATTAACCTAATCTT 736

QY 21 LeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluArgAlaIleAspGluLeu 40

DB 737 CTGGGGAAATTCAGACAGACACCTGGCTTTCGTAATGAAGAGAGTGCTACAGCTGATTTG 796

QY 41 LysLysGlnAlaIleGluAspLysGluAlaThrThrAlaIleGluAlaAlaSerSerAsp 60

DB 797 ACACGACAGCGGTAGCGGATCTACTGTGCAGCAGCGGAGCTGAAATGCTGGGGCAGCA 856

QY 61 AlaLeuGluAlaLeuAla 66

DB 857 GCTTGGGAAGCAGCGGCA 874

RESULT 13

AAQ06019

ID AAQ06019 standard; DNA; 2383 BP.

XX AAQ06019;

DT 17-JAN-1991 (first entry)

XX Sequence encoding Protein G with three active sites.

XX Immunoglobulin; ds.

OS Streptococcus sp. Lancefield Group G strain.

XX Key Location/Qualifiers

FT CDS 1..2355

FT /\*tag= a

XX US4956296-A.

XX 11-SEP-1990.

XX 20-JUN-1988; 88US-0209236.

XX 20-JUN-1988; 88US-0209236.

XX 14-FEB-1986; 86US-0829354.

XX 23-APR-1986; 86US-0854887.

XX 17-FEB-1987; 87WO-US00329.

XX 19-JUN-1987; 87US-0063959.

XX (GENE-) GENEX CORP.

XX Fahnestock SR;

XX WPI; 1990-297491/39.

DR P-PSDB; AAR07014.

XX Recombinant Protein G variants - obtd. using a cloned gene

PT encoding Protein G from Streptococcus sp., used for binding

XX immunoglobulin.

XX Disclosure; Fig 9; 48pp; English.

CC Fragments and variants of the sequence are claimed esp. where

CC incorporated into a non-pathogenic host eg. E.coli, and expressed

CC at high levels.

CC The variants have a higher binding efficiency and capacity for

CC immunoglobulin, and may be used for purifying, detecting and

CC isolating antibodies.

XX Sequence 2383 BP; 867 A; 411 C; 479 G; 626 T; 0 other;

Alignment Scores:

Pred. No.: 2.05e-12 Length: 2383

Score: 155.00 Matches: 37

Percent Similarity: 62.12% Conservative: 4

Best Local Similarity: 56.06% Mismatches: 25

Query Match: 19.95% Indels: 0

DB: 11 Gaps: 0

US-09-847-539A-6 (1-159) x AAQ06019 (1-2383)

QY 1 ValAspSerProIleGluGlnProArgIleIleProAsnGlyGlyThrLeuThrAsnLeu 20

DB 676 GTTGATTCCACCAATCGAAGATACCCCAATTATTGCGTAATGGTGGTGAATTAACCTAATCTT 735

QY 21 LeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluArgAlaIleAspGluLeu 40

DB 736 CTGGGGAAATTCAGACAGACACCTGGCTTTCGTAATGAAGAGAGTGCTACAGCTGATTTG 795

QY 41 LysLysGlnAlaIleGluAspLysGluAlaThrThrAlaIleGluAlaAlaSerSerAsp 60

DB 796 ACACGACAGCGGTAGCGGATCTACTGTGCAGCAGCGGAGCTGAAATGCTGGGGCAGCA 855

QY 61 AlaLeuGluAlaLeuAla 66

DB 856 GCTTGGGAAGCAGCGGCA 873

RESULT 14

AAQ10002

ID AAQ10002 standard; DNA; 2383 BP.

XX AAQ10002;

DT 13-MAR-1991 (first entry)

XX Sequence encoding Streptococcus GX7805 protein G.

XX Immunoglobulins; Ig; ds.

OS Streptococcus sp GX7805.

XX Key Location/Qualifiers

FT CDS 577..2355

FT /\*tag= a

XX US4977247-A.

XX 11-DEC-1990.

XX 19-MAY-1989; 89US-0354264.

XX 19-MAY-1989; 89US-0354264.

XX 14-FEB-1986; 86US-0829354.

XX 23-APR-1986; 86US-0854887.

XX 17-FEB-1987; 87WO-US00329.

XX 19-JUN-1987; 87US-0063959.

XX 20-JUN-1988; 88US-0209236.



```
Db 677 GTTGATTCACCAATCGAAGATACCCCAATTATTTCGTAATGGTGGTGAATTAACATACTTT 736
QY 21 LeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaIleAspGluLeu 40
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 737 CTGGGGAATTACAGACACACTGGCTTTCGTAATGAAGAGAGTCTACAGCTGATTG 796
QY 41 LysLysGlnAlaIleGluAspLysGluAlaThrAlaIleGluAlaAlaSerSerAsp 60
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 797 ACAGCAGCAGCGTACCCATACCTGTGGCAGCAGCGCAGCTGAAATGCTGGGGCAGCA 856
QY 61 AlaLeuGluAlaLeuAla 66
    ||||| ||||| |||||
Db 857 GCTTGGGAAGCAGCGGCA 874

RESULT 9
AAN91094
ID AAN91094 standard; DNA; 1950 BP.
AC AAN91094;
XX
DT 04-JUL-1990 (first entry)
XX
DE Protein G gene.
XX
KW Protein G; immunoglobulin; Fc receptor; ds.
XX
OS Streptococcus sp.
XX
FH Key Location/Qualifiers
FT -35_signal 465..470
FT /*tag= c
FT -10_signal 487..492
FT /*tag= b
FT RBS 565..571
FT /*tag= d
FT CDS 578..1918
FT /*tag= a
XX
PN WO8810306-A.
XX
XX
PD 29-DEC-1988.
XX
PF 20-JUN-1988; 88WO-US02084.
XX
PR 19-JUN-1987; 87US-0063959.
XX
PA (GENE-) GENEX CORP.
XX
PI Fahnstock SR;
XX
DR WPI: 1989-023848/03.
DR P-PSDB; AAP95030.
XX
PT Cloned protein G variant genes -
PT expressing proteins having immunoglobulin-binding properties of
PT protein G and derived from Streptococcus sp.
XX
PS Disclosure; ; 116pp; English.
XX
CC Gene encodes protein G of non-pathogenic streptococcus sp. allowing
CC isolation of the protein and variants, useful as bacterial Fc receptors
CC eg in purification and detection of Abs., screening of hybridoma clones
CC and treatment of disease.
XX
SQ Sequence 1950 BP; 706 A; 324 C; 399 G; 521 T; 0 other;

Alignment Scores:
Pred. No.: 1.57e-12 Length: 1950
Score: 155.00 Matches: 37
Percent Similarity: 62.1% Conservative: 4
Best Local Similarity: 56.0% Mismatches: 25
Query Match: 19.95% Indels: 0
DB: 10 Gaps: 0
```

```
US-09-847-539A-6 (1-159) x AAN91094 (1-1950)
QY 1 ValAspSerProIleGluGlnProArgIleIleProAsnGlyGlyThrLeuThrAsnLeu 20
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 677 GTTGATTCACCAATCGAAGATACCCCAATTATTTCGTAATGGTGGTGAATTAACATACTTT 736
QY 21 LeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaIleAspGluLeu 40
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 737 CTGGGGAATTACAGACACACTGGCTTTCGTAATGAAGAGAGTCTACAGCGGATTG 796
QY 41 LysLysGlnAlaIleGluAspLysGluAlaThrAlaIleGluAlaAlaSerSerAsp 60
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 797 ACAGCAGCAGCGTACCCATACCTGTGGCAGCAGCGCAGCTGAAATGCTGGGGCAGCA 856
QY 61 AlaLeuGluAlaLeuAla 66
    ||||| ||||| |||||
Db 857 GCTTGGGAAGCAGCGGCA 874

RESULT 10
AAQ06017
ID AAQ06017 standard; DNA; 1950 BP.
XX
AC AAQ06017;
XX
DT 17-JAN-1991 (first entry)
XX
DE Sequence encoding Protein G.
XX
KW Immunoglobulin; ds.
XX
OS Streptococcus sp. Lancefield Group G strain.
XX
FH Key Location/Qualifiers
FT CDS 578..1921
FT /*tag= a
XX
PN US4956296-A.
XX
XX
PD 11-SEP-1990.
XX
PF 20-JUN-1988; 88US-0209236.
XX
PR 20-JUN-1988; 88US-0209236.
PR 14-FEB-1986; 86US-0829354.
PR 23-APR-1986; 86US-0854887.
PR 17-FEB-1987; 87WO-US00329.
PR 19-JUN-1987; 87US-0063959.
XX
PA (GENE-) GENEX CORP.
XX
PI Fahnstock SR;
XX
DR WPI: 1990-297491/39.
DR P-PSDB; AAR07012.
XX
PT Recombinant Protein G variants - obtd. using a cloned gene
PT encoding Protein G from Streptococcus sp.; used for binding
PT immunoglobulin.
XX
PS Disclosure; Fig 3a-c; 48pp; English.
XX
CC Fragments and variants of the sequence are claimed esp. where
CC incorporated into a non-pathogenic host eg. E.coli, and expressed
CC at high levels.
CC The variants have a higher binding efficiency and capacity for
CC immunoglobulin, and may be used for purifying, detecting and
CC isolating antibodies.
XX
SQ Sequence 1950 BP; 706 A; 323 C; 398 G; 523 T; 0 other;

Alignment Scores:
Pred. No.: 1.57e-12 Length: 1950
```



```

ID AAQ06018 standard; DNA; 1949 BP.
XX
AC AAQ06018;
XX
DT 17-JAN-1991 (first entry)
XX
DE Sequence encoding Protein G with two active sites.
XX
KW Immunoglobulin; ds.
XX
OS Streptococcus sp. Lancefield Group G strain.
XX
FH Key Location/Qualifiers
FT CDS 577..1920
FT FT /*tag= a
FT FT -35_signal 465..470
FT FT /*tag= b
FT FT -10_signal 487..492
FT FT /*tag= c
FT FT RBS 565..571
FT FT /*tag= d
XX
XX USA956296-A.
XX
XX 11-SEP-1990.
XX
XX 20-JUN-1988; 88US-0209236.
XX
XX 20-JUN-1988; 88US-0209236.
XX
XX 14-FEB-1986; 86US-0829354.
XX
XX 23-APR-1986; 86US-0854887.
XX
XX 17-FEB-1987; 87WO-US00329.
XX
XX 19-JUN-1987; 87US-0063959.
XX
XX (GENE-) GENEX CORP.
XX
XX Fahnstock SR;
XX
XX WPI; 1990-297491/39.
XX
XX P-PSDB; AAR07013.
XX
XX Recombinant Protein G variants - obtd. using a cloned gene
XX encoding Protein G from Streptococcus sp., used for binding
XX immunoglobulin.
XX
XX Disclosure; Fig 8a-c; 48pp; English.
XX
XX Fragments and variants of the sequence are claimed esp. where
XX incorporated into a non-pathogenic host eg. E.coli, and expressed
XX at high levels.
XX
XX The variants have a higher binding efficiency and capacity for
XX immunoglobulin, and may be used for purifying, detecting and
XX isolating antibodies.
XX
XX Sequence 1949 BP; 705 A; 323 C; 398 G; 523 T; 0 other;

Alignment Scores:
Pred. No.: 1.57e-12 Length: 1949
Score: 155.00 Matches: 37
Percent Similarity: 62.12% Conservatives: 4
Best Local Similarity: 56.06% Mismatches: 25
Query Match: 19.95% Indels: 0
DB: 11 Gaps: 0

US-09-847-539A-6 (1-159) x AAQ06018 (1-1949)
Qy 1 ValAspSerProIleGluGlnProArgIleIleProAsnGlyThrLeuThrAsnLeu 20
Db 676 GTTGATTCCACCAATGAGAGACCAATATTCTGTAATGGTGAATTAACATACTT 735
Qy 21 LeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaIleAspGluLeu 40
Db 736 CTGGGAATTCAGAGACAACACTGGCTTTGGTAAATGAAGAGAGTGCTACAGCTGATTG 795

QY 41 LysLysGlnAlaIleGluAspLysGluAlaThrThrAlaIleGluAlaAlaSerSerAsp 60
Db 796 ACAGCACCGGTAGCGGATAGTGTGGCAGCAGCGGCTGAAATGCTGGGGCAGCA 855
QY 61 AlaLeuGluAlaLeuAla 66
Db 856 GCTTGGGAAGCAGCGGCA 873

RESULT 8
AAN91093
ID AAN91093 standard; DNA; 1950 BP.
XX
XX AAN91093;
XX
XX 04-JUL-1990 (first entry)
XX
XX Protein G gene.
XX
XX Protein G; Immunoglobulin; Fc receptor; ds.
XX
XX Streptococcus sp.
XX
XX Key Location/Qualifiers
XX -35_signal 465..470
XX FT /*tag= c
XX FT -10_signal 487..492
XX FT /*tag= b
XX FT RBS 565..571
XX FT /*tag= d
XX FT CDS 578..1918
XX FT /*tag= a
XX
XX WO8810306-A.
XX
XX 29-DEC-1988.
XX
XX 20-JUN-1988; 88WO-US02084.
XX
XX 19-JUN-1987; 87US-0063959.
XX
XX (GENE-) GENEX CORP.
XX
XX Fahnstock SR;
XX
XX WPI; 1989-023848/03.
XX
XX P-PSDB; AAP95030.
XX
XX Cloned protein G variant genes -
XX expressing proteins having immunoglobulin-binding properties of
XX protein G and derived from Streptococcus sp.
XX
XX Disclosure; ; lpp; English.
XX
XX Gene encodes protein G of non-pathogenic streptococcus sp. allowing
XX isolation of the protein and variants, useful as bacterial Fc receptors
XX eg in purification and detection of Abs., screening of hybridoma clones
XX and treatment of disease.
XX
XX Sequence 1950 BP; 705 A; 323 C; 398 G; 524 T; 0 other;

Alignment Scores:
Pred. No.: 1.57e-12 Length: 1950
Score: 155.00 Matches: 37
Percent Similarity: 62.12% Conservatives: 4
Best Local Similarity: 56.06% Mismatches: 25
Query Match: 19.95% Indels: 0
DB: 10 Gaps: 0

US-09-847-539A-6 (1-159) x AAN91093 (1-1950)
Qy 1 ValAspSerProIleGluGlnProArgIleIleProAsnGlyThrLeuThrAsnLeu 20
Db 1 ValAspSerProIleGluGlnProArgIleIleProAsnGlyThrLeuThrAsnLeu 20

```

CC protein G of group G Streptococcus. GRAB protein and peptides derived  
 CC from it are used in vaccine compositions for generating a protective  
 CC immune response against group A Streptococcus. Antibodies against GRAB  
 CC are useful for treating Streptococcus pyogenes infections. The protein  
 CC is also useful for purifying alpha2M from a sample. The present sequence  
 CC is a DNA encoding partial GRAB protein from S. pyogenes strain AP49.  
 CC The protein has alpha2M binding region and is useful in vaccine  
 CC composition.  
 XX

SQ Sequence 853 BP; 295 A; 171 C; 197 G; 190 T; 0 other;

Alignment Scores:  
 Pred. No.: 5,8e-69 Length: 853  
 Score: 536.00 Matches: 113  
 Percent Similarity: 94.2% Conservative: 2  
 Best Local Similarity: 92.62% Mismatches: 7  
 Query Match: 68.98% Indels: 0  
 DB: 21 Gaps: 0

US-09-847-539A-6 (1-159) x AAD00563 (1-853)

QY 1 ValAspSerProIleGluGlnProArgIleIleProAsnGlyClyThrLeuThrAsnLeu 20  
 DB 46 GTTGACTCACTATCGACAGCTCGAATTATTCCAAATGGGGAACCTTAATCTT 105  
 QY 21 LeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaIleAspGluLeu 40  
 DB 106 CTTGGCAATGCTCCAGAAACTGGCATTTACGTAATGAAGAAGCCATTGATGAATTA 165  
 QY 41 LysLysGlnAlaIleGluAspLysGluAlaThrThrAlaIleGluAlaIleAspGluLeu 60  
 DB 166 AAAAAACAAGCTATTGAGGATAAAGAAGCTACGACAGCTATAGAAGCAGCAAGTTCCAGAT 225  
 QY 61 AlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaIleValVal 80  
 DB 226 GCCTTAGAAGCATTTAGCGGATCAAGCAGACGCTTTACAAATCAGAAGAAGCTGCAGTAGTT 285  
 QY 81 LysAlaAspAsnAlaAlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeu 100  
 DB 286 CAATCAGATACGCTCTAGTACGCTTACGAGCCTTAGCAGCATTTAGCGGATCAAGCAGCAGCTTTA 345  
 QY 101 GlnSerGluGluAlaGluValValGlnSerAspAsnAlaAlaSerAspAlaTrpGluLys 120  
 DB 346 CAATCAGAAGAAGCTCAGTAGTTTCAATCAGATAGCGTCTGGTGCAGCGCTTAGAAGCA 405  
 QY 121 AlaAla 122  
 DB 406 TTGGCG 411

RESULT 6  
 AAD00561

ID AAD00561 standard; DNA; 777 BP.

AC AAD00561;

DT 29-AUG-2000 (first entry)

DE Streptococcus pyogenes strain KTL9 partial GRAB protein encoding DNA.

KW GRAB protein; protein G related alpha2M binding protein; vaccine;  
 KW alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;  
 KW immune response; Streptococcus pyogenes infection; ds.

OS Streptococcus pyogenes.

XX Key Location/Qualifiers  
 FT CDS 1..777

FT /\*tag= a  
 FT /product= "GRAB protein"  
 FT /note= "Does not include stop codon"  
 FT /partial

PN W0200026240-A2.

XX

PD 11-MAY-2000.

XX 02-NOV-1999; 99WO-GB03631.

XX 02-NOV-1998; 98GB-0023975.

XX (ACTI-) ACTINOVA LTD.

XX Bjorck LH, Rasmussen M;

PI WPI: 2000-365572/31.

DR P-PSDB; AAY71043.

XX New alpha2M binding protein for generating a protective immune response  
 PT to group A streptococcus and purifying the binding protein -

XX Claim 13; Page 63-64; 67pp; English.

XX The patent discloses a new family of proteins termed GRAB (protein G  
 CC related alpha2M binding protein) from Streptococcus pyogenes which have  
 CC the ability to bind alpha2-macroglobulin (alpha2M) and show homology to  
 CC protein G of group G Streptococcus. GRAB protein and peptides derived  
 CC from it are used in vaccine compositions for generating a protective  
 CC immune response against group A Streptococcus. Antibodies against GRAB  
 CC are useful for treating Streptococcus pyogenes infections. The protein  
 CC is also useful for purifying alpha2M from a sample. The present sequence  
 CC is a DNA encoding partial GRAB protein from S. pyogenes strain KTL9.  
 CC The protein has alpha2M binding region and is useful in vaccine  
 CC composition.

SQ Sequence 777 BP; 269 A; 151 C; 178 G; 179 T; 0 other;

Alignment Scores:

Pred. No.: 1.01e-68 Length: 777  
 Score: 534.00 Matches: 112  
 Percent Similarity: 95.08% Conservative: 4  
 Best Local Similarity: 91.80% Mismatches: 6  
 Query Match: 68.73% Indels: 0  
 DB: 21 Gaps: 0

US-09-847-539A-6 (1-159) x AAD00561 (1-777)

QY 1 ValAspSerProIleGluGlnProArgIleIleProAsnGlyClyThrLeuThrAsnLeu 20  
 DB 61 GTTGACTCACTATCGACAGCTCGAATTATTCCAAATGGGGAACCTTAATCTT 120

QY 21 LeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaIleAspGluLeu 40  
 DB 121 CTTGGCAATGCTCCAGAAACTGGCATTTACGTAATGAAGAAGGCCATTGATGAATTA 180

QY 41 LysLysGlnAlaIleGluAspLysGluAlaThrThrAlaIleGluAlaIleAspGluLeu 60  
 DB 181 AAAAAACAAGCTATTGAGGATAAAGAAGCTACGACAGCTATAGAAGCAGCAAGTTCCAGAT 240

QY 61 AlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaIleValVal 80  
 DB 241 GCCTTAGAAGCATTTAGCGGATCAAGCAGACGCTTTTCAATCAGAAGAAGCTGCAGTAGTT 300

QY 81 LysAlaAspAsnAlaAlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeu 100  
 DB 301 CAATCAGATACGCTCTAGTACGCTTGAAGAAGCATTTGGCGGATCAACAGACGCTTTA 360

QY 101 GlnSerGluGluAlaGluValValGlnSerAspAsnAlaAlaSerAspAlaTrpGluLys 120  
 DB 361 CAATCAGAAGAAGCTCGCGTTGTTAAAGCGGATAACGCTGCTAGTCACACTTTTAGAAGCA 420

QY 121 AlaAla 122

DB 421 TTGGCG 426

RESULT 7  
 AAD00518

Db 454 AAAGAGAAAGACAAAACGTTAATACCTTCTTACAACTGGTGAAGAG 501  
|||||

RESULT 4

AAD00562

ID AAD00562 standard; DNA; 469 BP.

XX AC AAD00562;

XX 29-AUG-2000 (first entry)

XX DE Streptococcus pyogenes strain AP1 partial GRAB protein encoding DNA.

XX DE GRAB protein; protein G related alpha2M binding protein; vaccine;

KW alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;

KW immune response; Streptococcus pyogenes infection; ds.

XX OS Streptococcus pyogenes.

XX FH Key Location/Qualifiers

FT CDS 1..468

FT /\*tag= a

FT /product= "GRAB protein"

FT /partial

XX XX WO200026240-A2.

XX XX 11-MAY-2000.

XX XX 02-NOV-1999; 99WO-GB03631.

XX XX 02-NOV-1998; 98GB-0023975.

XX XX (ACTI-) ACTINOVA LTD.

XX XX Bjorck LH, Rasmussen M;

XX XX WPI; 2000-365572/31.

DR P-PSDB; AAY71044.

XX XX New alpha2M binding protein for generating a protective immune response

PT to group A streptococcus and purifying the binding protein -

XX XX Claim 13; Page 64; 67pp; English.

XX XX The patent discloses a new family of proteins termed GRAB (protein G

CC related alpha2M binding protein) from Streptococcus pyogenes which have

CC the ability to bind alpha2-macroglobulin (alpha2M) and show homology to

CC protein G of group A Streptococcus. GRAB protein and peptides derived

CC from it are used in vaccine compositions for generating a protective

CC immune response against group A Streptococcus. Antibodies against GRAB

CC are useful for treating Streptococcus pyogenes infections. The protein

CC is also useful for purifying alpha2M from a sample. The present sequence

CC is a DNA encoding partial GRAB protein from S. pyogenes strain API.

CC The protein has alpha2M binding region and is useful in vaccine

CC composition.

XX XX SQ Sequence 469 BP; 180 A; 90 C; 99 G; 100 T; 0 other;

#### Alignment Scores:

Pred. No.: 9.74e-101 Length: 469  
Score: 750.00 Matches: 154  
Percent Similarity: 99.35% Conservative: 0  
Best Local Similarity: 99.35% Mismatches: 1  
Query Match: 96.53% Indels: 0  
DB: 21 Gaps: 0

US-09-847-539A-6 (1-159) x AAD00562 (1-469)

Qy 2 AspSerProIleGluInProArgIleProAsnGlyGlyThrLeuThrAsnLeuLeu 21

|||||

Db 1 GACTCACCTATCGACAGCCCTAGATTTATTCCTAAATGGCGGAACCTTAAATTAATCTCTT 60

QY 22 GlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaIleAspGluLeuLys 41  
|||||  
Db 61 GGCAATGCTCCAGAAAACTGGCATTTACGTAATGAAGAAAGAGCCATTGATGAATTTAAA 120  
QY 42 LysGlnAlaIleGluAspLysGluAlaThrThrAlaIleGluAlaAlaSerSerAspAla 61  
|||||  
Db 121 AAACAAGCTATTGAGGATAAGGAAGCTACGACAGCTATAGAAGCAGCAAGTTCAGATGCC 180  
QY 62 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLys 81  
|||||  
Db 181 TTAGAACCATTAGCGGATCAACAGACGCTTTACAATCAGAAGAAGCTGCGGTGTTTAAA 240  
QY 82 AlaAspAsnAlaAlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGln 101  
|||||  
Db 241 CGCGATAACGCTGCTAGTAGCGCTTTAGAACGATTTGGCGGATCAACAGAGAGCGCTTTACAA 300  
QY 102 SerGluGluAlaGluValValGlnSerAspAsnAlaAlaSerAspAlaTtpGluLysAla 121  
|||||  
Db 301 TCAGAAGAAGCTGAAGTAGTTCAATCAGATACGCTGCTAGTGACGCTGGGAAAAAGCA 360  
QY 122 AlaThrProIleAlaLeuAspValLysThrLysAspThrLysProValValLysLys 141  
|||||  
Db 361 GCRACTCCATCGCTTTAGATGTTAAGAAAACTAAAGATACAAAAACCTGTAGTTAAAAAA 420  
QY 142 GluGluArgGlnAsnValAsnThrLeuProThrThrGlyGluGlu 156  
|||||  
Db 421 GAAGAAAGACAAAAACGTTAATACCTTCTTACAACTGGTGAAGAG 465

#### RESULT 5

AAD00563

ID AAD00563 standard; DNA; 853 BP.

XX AC AAD00563;

XX XX 29-AUG-2000 (first entry)

XX XX Streptococcus pyogenes strain AP49 partial GRAB protein encoding DNA.

XX KW GRAB protein; protein G related alpha2M binding protein; vaccine;

KW alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;

KW immune response; Streptococcus pyogenes infection; ds.

XX OS Streptococcus pyogenes.

XX FH Key Location/Qualifiers

FT CDS 37..852

FT /\*tag= a

FT /product= "GRAB protein"

FT /partial

XX XX WO200026240-A2.

XX XX 11-MAY-2000.

XX XX 02-NOV-1999; 99WO-GB03631.

XX XX 02-NOV-1998; 98GB-0023975.

XX XX (ACTI-) ACTINOVA LTD.

XX XX Bjorck LH, Rasmussen M;

XX XX WPI; 2000-365572/31.

DR P-PSDB; AAY71045.

XX XX New alpha2M binding protein for generating a protective immune response

PT to group A streptococcus and purifying the binding protein -

XX XX Claim 13; Page 64; 67pp; English.

XX XX The patent discloses a new family of proteins termed GRAB (protein G

CC related alpha2M binding protein) from Streptococcus pyogenes which have

CC the ability to bind alpha2-macroglobulin (alpha2M) and show homology to

CC protein G of group A Streptococcus. GRAB protein and peptides derived

CC from it are used in vaccine compositions for generating a protective

CC immune response against group A Streptococcus. Antibodies against GRAB

CC are useful for treating Streptococcus pyogenes infections. The protein

CC is also useful for purifying alpha2M from a sample. The present sequence

CC is a DNA encoding partial GRAB protein from S. pyogenes strain API.

CC The protein has alpha2M binding region and is useful in vaccine

CC composition.

XX XX SQ Sequence 469 BP; 180 A; 90 C; 99 G; 100 T; 0 other;

CC protein G of group G Streptococcus. GRAB protein and peptides derived  
CC from it are used in vaccine compositions for generating a protective  
CC immune response against group A Streptococcus. Antibodies against GRAB  
CC are useful for treating Streptococcus pyogenes infections. The protein  
CC is also useful for purifying alpha2M from a sample. The present sequence  
CC is a DNA encoding full-length GRAB protein from S. pyogenes strain SF370.  
XX  
SQ Sequence 764 BP; 279 A; 131 C; 159 G; 195 T; 0 other;

Alignment Scores:  
Pred. No.: 2.03e-104 Length: 764  
Score: 777.00 Matches: 159  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 21 Gaps: 0

US-09-847-539A-6 (1-159) x AAD00559 (1-764)

QY 1 ValAspSerProIleGluGlnProArgIleIleProAsnGlyGlyThrLeuThrAsnLeu 20  
DB 159 GTTGACTCACTTACGAGCAGCTCGAATATTCCAAATGGCGGAACCTTAATCTT 218  
QY 21 LeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaIleAspGluLeu 40  
DB 219 CTGGCAATGCTCCAGAAAACCTGGCATTCAGTATGAGAAGAGCCATTGATGATTA 278  
QY 41 LysLysGlnAlaIleGluAspLysGluAlaThrAlaIleGluAlaAlaSerSerAsp 60  
DB 279 AAAAAACAAGCTATTGAGATAAAGAAAGCTACGACAGCTATAGACGAGCAAGTTCCAGAT 338  
QY 61 AlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluAlaAlaValVal 80  
DB 339 GCGTTAGAAGCATTCGCGGATCAACAGACGCTTTCACAAATCAGAGAAGCTGCGGTGTT 398  
QY 81 LysAlaAspAsnAlaAlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeu 100  
DB 399 AAGCGGATAACGCTGCTAGTCAGCGCTTGAAGAGCATTCGCGGATCAACAGACGCTTAA 458  
QY 101 GlnSerGluGluAlaGluValValGlnSerAspAsnAlaAlaSerAspAlaIleAspGluLys 120  
DB 459 CAATCAGAGAAGAGCTGAAGTACTTCAATCAGATAACGCTGCTAGTGACGCTGGGAAAAA 518  
QY 121 AlaAlaThrProIleAlaLeuAspValLysLysThrLysAspThrLysProValValLys 140  
DB 519 GCAGCACTCCCAATCGCTTAGATGTTAAGAAACCTTAAAGATACAAACCTGTAGTTAA 578  
QY 141 LysGluGluArgGlnAsnValAsnThrLeuProThrThrGlyGluGluSerAsnPro 159  
DB 579 AAGAAGAAGAACAAACCGTTAATACCTTCTACAACTGGTGAGAGTCTAACCCA 635

#### RESULT 3

AAD00564  
ID AAD00564 standard; DNA; 504 BP.

XX AC AAD00564;

XX DT 29-AUG-2000 (first entry)

XX DE Streptococcus pyogenes strain KTL3 partial GRAB protein encoding DNA.

XX KW GRAB protein; protein G related alpha2M binding protein; vaccine;

XX KW alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;

XX KW Immune response; Streptococcus pyogenes infection; ds.

XX OS Streptococcus pyogenes.

XX PH Key Location/Qualifiers

XX FT CDS

XX FT 1..504

XX FT /\*tag= a

XX FT /product= \*GRAB protein\*

XX FT /partial

PN WO200026240-A2.  
XX  
PD 11-MAY-2000.  
XX  
PF 02-NOV-1999; 99WO-GB03631.  
XX  
PR 02-NOV-1998; 98GB-0023975.  
XX  
PA (ACTI-) ACTINOVA LTD.  
XX  
PI Bjorck LH, Rasmussen M;  
XX  
DR WPI: 2000-365572/31.  
DR P-PSDB; AAY71046.  
XX  
PT New alpha2M binding protein for generating a protective immune response  
PT to group A streptococcus and purifying the binding protein -  
XX  
PS Claim 13; Page 65; 67pp; English.

CC The patent discloses a new family of proteins termed GRAB (protein G  
CC related alpha2M binding protein) from Streptococcus pyogenes which have  
CC the ability to bind alpha2-macroglobulin (alpha2M) and show homology to  
CC protein G of group G Streptococcus. GRAB protein and peptides derived  
CC from it are used in vaccine compositions for generating a protective  
CC immune response against group A Streptococcus. Antibodies against GRAB  
CC are useful for treating Streptococcus pyogenes infections. The protein  
CC is also useful for purifying alpha2M from a sample. The present sequence  
CC is a DNA encoding partial GRAB protein from S. pyogenes strain KTL3.  
CC The protein has alpha2M binding region and is useful in vaccine  
CC composition.

SQ Sequence 504 BP; 188 A; 97 C; 108 G; 111 T; 0 other;

#### Alignment Scores:

Pred. No.: 3.65e-102 Length: 504  
Score: 760.00 Matches: 156  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 97.81% Indels: 0  
DB: 21 Gaps: 0

US-09-847-539A-6 (1-159) x AAD00564 (1-504)

QY 1 ValAspSerProIleGluGlnProArgIleIleProAsnGlyGlyThrLeuThrAsnLeu 20  
DB 34 GTTGACTCACTTACGAGCAGCTCGAATATTCCAAATGGCGGAACCTTAATCTT 93  
QY 21 LeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaIleAspGluLeu 40  
DB 94 CTGGCAATGCTCCAGAAAACCTGGCATTCAGTATGAGAAGAGCCATTGATGATTA 153  
QY 41 LysLysGlnAlaIleGluAspLysGluAlaThrAlaIleGluAlaAlaSerSerAsp 60  
DB 154 AAAAAACAAGCTATTGAGATAAAGAAAGCTACGACAGCTATAGACGAGCAAGTTCCAGAT 213  
QY 61 AlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVal 80  
DB 214 GCGTTAGAAGCATTCGCGGATCAACAGACGCTTTCACAAATCAGAGAAGCTGCGGTGTT 273  
QY 81 LysAlaAspAsnAlaAlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeu 100  
DB 274 AAGCGGATAACGCTGCTAGTCAGCGCTTGAAGAGCATTCGCGGATCAACAGACGCTTAA 333  
QY 101 GlnSerGluGluAlaGluValValGlnSerAspAsnAlaAlaSerAspAlaIleAspGluLys 120  
DB 334 CAATCAGAGAAGAGCTGAAGTACTTCAATCAGATAACGCTGCTAGTGACGCTGGGAAAAA 393  
QY 121 AlaAlaThrProIleAlaLeuAspValLysLysThrLysAspThrLysProValValLys 140  
DB 394 GCAGCACTCCCAATCGCTTAGATGTTAAGAAACCTTAAAGATACAAACCTGTAGTTAA 453  
QY 141 LysGluGluArgGlnAsnValAsnThrLeuProThrThrGlyGluGlu 156

FT	/*tag= a	/*product= "GRAB protein"
FT	1.99	
FT	slg_peptide	/*tag= b
FT	mat_peptide	100..651
FT		/*tag= c
FT		/*product= "Mature GRAB protein"

PN WO200026240-A2.

11-MAY-2000.

AA 02-NOV-1999; 99WO-GB03631.

AA 02-NOV-1998; 98GB-0023975.

PA (ACTI-) ACTINOVA LTD.

PI Bjorck LH, Rasmussen M;

WPI; 2000-365572/31.

UR P-PSDB; AAI/1042.  
XX

PI new alpha2m binding protein for generating a protective immune response

PT to group A streptococcus and purifying the binding protein -

PS Claim 13; Page 63; 67pp; English.

AA The patent discloses a new family

related alpha2m binding protein) from *Streptococcus pyogenes* which have the ability to bind alpha2-macroglobulin (alpha2m) and show homology to protein G of group G *Streptococcus*. GRAB protein and peptides derived from it are used in vaccine compositions for generating a protective immune response against group A *Streptococcus*. Antibodies against GRAB are useful for treating *Streptococcus pyogenes* infections. The protein is also useful for purifying alpha2m from a sample. The present sequence is a DNA encoding full-length GRAB protein from *S. pyogenes* strain 3F370.

Sequence 654 BP; 234 A; 118 C; 143 G; 159 T; 0 other;

**Alignment Scores:**

Pred. No.:	1,65e-104	Length:	654
Score:	77.0%	Matches:	159
Percent Similarity:	100.0%	Conservative:	0
Best Local Similarity:	100.0%	Mismatches:	0
Query Match:	100.0%	Indels:	0
DB:	21	Gaps:	0

US-09-847-539A-6 (1-159) x AAD00560 (1-654)

QY 1 ValAspSerProIleGluGlnProArgIleIleProAsnGlyGlyThrLeuThrAsnLeu 20

Db 100 GTGACTCACCTATCGAACAGCCTCGAATTATTCCAAATGCGGGAACCTTAACCTAATCTT 159

QY 21 LeuGlyAsnAlaProGluLysLeuAlaLeuArqAsnGluGluArqAlaIleAspGluLeu 40

Db 160 CTTGGCAATGCTCCAGAAAACCTGGCATTACGTAATGAAGAAAGAGCCATTGATGAATTA 219

Qv 41 LysLysGlnAlaIleGluAspLysGluAlaThrThrAlaIleGluAlaAlaSerSerAsp 60

220 AAAAAACAAGCTATTGAGGATAAAGAAGCTACGACAGCTATAGAACGACGAAGTTCAGAT 279

0y 61 A]a[euG]uA]a[euA]aAsG]nThraSsA]a[euG]nSerG]uG]uA]aA]aVa]80

db 280 GCCTTAGAAGCATTAGCGGATCAAAACAGACCGTTTACATACACACACAGCCCTTCTT 339

[illegible][illegible][illegible]

1000

GenCore version 5.1.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 13, 2002, 02:10:56 : Search time 229.572 Seconds  
(without alignments)  
1189.122 Million cell updates/sec

Title: US-09-847-539A-6

Perfect score: 777

Sequence: 1 VDSPIEQPRIIPNGTTLNL.....KKEERQNVNLTPTTGESNP 159

Scoring table: BLOSUM62

Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters: -DEV=xlh

-Q=/cgn2\_1/USPTO\_spool/JUS09847539/runat\_10102002\_092548\_3413/app\_query.fasta\_1.526  
-DB=N\_Geneseq\_032802 -OFMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=40  
-MODE=LOCAL -OUTFMT=pto -NORW=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09847539@cgn\_1\_1.48 -runat\_10102002\_092548\_3413 -NCPU=6 -ICPU=3  
-NO\_XLPHY -NO\_MMAPP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FPGAPOP=6 -FPGAPEXT=7  
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : N\_Geneseq\_032802.\*  
1: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	777	100.0	654	21	AAD00560	Streptococcus pyog
2	777	100.0	764	21	AAD00559	Streptococcus pyog
3	760	97.8	504	21	AAD00564	Streptococcus pyog
4	750	96.5	469	21	AAD00562	Streptococcus pyog
5	536	69.0	853	21	AAD00563	Streptococcus pyog
6	534	68.7	777	21	AAD00561	Streptococcus pyog
7	155	19.9	1949	11	AAQ06018	Sequence encoding
8	155	19.9	1950	10	AAQ06019	Protein G gene. S
9	155	19.9	1950	10	AAQ06017	Protein G gene. S
10	155	19.9	1950	11	AAQ06016	Sequence encoding
11	155	19.9	1950	12	AAQ10001	Sequence encoding
12	155	19.9	1950	15	AAQ64644	Streptococcus prot
13	155	19.9	2383	11	AAQ06019	Sequence encoding
14	155	19.9	2383	12	AAQ10002	Sequence encoding
15	155	19.9	2383	15	AAQ75036	Streptococcus prot
16	142	18.3	1576	21	AAZ47931	Streptococcus stra
17	105	13.5	798	10	AAQ91099	Protein G gene var
18	105	13.5	798	11	AAQ06009	Sequence encoding
19	105	13.5	798	12	AAQ10007	Type 4 GX7809 prot
20	105	13.5	798	15	AAQ64648	IGG-binding Strept
21	93	12.0	1380	23	AA584496	DNA encoding novel
22	93	12.0	1383	23	AA580225	DNA encoding novel
23	91	11.7	1555	16	AAQ86080	S. dysgalactiae ma
24	91	11.7	1555	16	AAQ89197	S. dysgalactiae ma
25	90	11.6	1469	8	AAQ70757	Sequence encoding
26	90	11.6	2121	23	ABL20337	Drosophila melanog
27	90	11.6	1145	23	ABL20336	Drosophila melanog
28	86	11.1	3411	23	AA582434	DNA encoding novel
29	85	10.9	2526	16	AAQ89196	S. dysgalactiae mi
30	84	10.8	2129	23	ABL03677	Drosophila melanog
31	84	10.8	5366	23	ABL03676	Drosophila melanog
32	83	10.7	141	16	AAQ84884	Expression cassett
33	83	10.7	603	16	AAQ84887	Chimeric syntheiti
34	83	10.7	627	16	AAQ84886	Synthetic protein
35	83	10.7	812	21	AAC38790	Arabidopsis thalia
36	83	10.7	2561	23	AA582435	DNA encoding novel
37	81	10.4	2010	24	AA518114	Streptococcus dysg
38	81	10.4	3069	23	AA577362	DNA encoding novel
39	81	10.4	3069	23	AA589837	DNA encoding novel
40	77	9.9	782	23	AA574086	DNA encoding novel
41	77	9.9	1779	23	AA579794	DNA encoding novel
42	76	9.8	441	22	AAH52591	S. epidermidis ope
43	76	9.8	3600	22	AAH54801	S. epidermidis gen
44	76	9.8	10136	17	AAT34578	Kinetochore protei
45	75	9.7	1290	22	AAH90726	CPE 30 coding sequ

ALIGNMENTS

RESULT 1

AAD00560

ID AAD00560 standard; DNA; 654 BP.

XX

AC AAD00560;

XX

DT 29-AUG-2000 (first entry)

XX

DE Streptococcus pyogenes strain SF370 GRAB protein coding region.

XX

KW GRAB protein; protein G related alpha2M binding protein; vaccine;

KW alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;

KW immune response; Streptococcus pyogenes infection; ds.

XX

OS Streptococcus pyogenes.

XX

FH Key

FT CDS Location/Qualifiers

1..654

**THIS PAGE BLANK (uspro)**

; LIBRARY: LAMBDA GEM(TM)-11 LIBRARY OF KENNETH NARVA  
; CLONE: PS69DIA  
; FEATURE:  
; NAME/KEY: Protein  
; LOCATION: 1..395  
US-08-049-783-6

Query Match 31.8%; Score 41; DB 1; Length 395;  
Best Local Similarity 40.9%; Pred. No. 1.9e+02;  
Matches 9; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

QY 5 EALADQTDALQSEEAIVVKADN 26  
| | : | | : | : | : |  
Db 231 EELLKVVDDLKKELEAAIKAEK 252

Search completed: October 13, 2002, 04:50:07  
Job time : 5.34225 secs



; COUNTRY: USA  
; ZIP: 20004  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US93/01676A  
; FILING DATE: 19930226  
; CLASSIFICATION:  
; PRIORITY APPLICATION DATA:  
; APPLICATION NUMBER: 07/842,644  
; FILING DATE: 02-FEB-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cooper, Iver P.  
; REGISTRATION NUMBER: 28,005  
; REFERENCE/DOCKET NUMBER: STAMSBROOK 1  
; TELEPHONE: 202-628-5197  
; TELEFAX: 202-737-3528  
; TELEX: 248633  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 370 amino acids  
; TYPE: AMINO ACID  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; PCT-US93-01676A-4

Query Match 31.8%; Score 41; DB 5; Length 370;  
Best Local Similarity 43.5%; Pred. No. 1.7e+02;  
Matches 10; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

QY 1 SDALALADOTDALQSEEAAYVK 23  
DB 343 ADSLMOLARQVSRLESQQAALPK 365

RESULT 39  
US-07-876-280-9  
; Sequence 9, Application US/07876280  
; Patent No. 5262158  
; GENERAL INFORMATION:  
; APPLICANT: Payne, Jewel M.  
; APPLICANT: Cannon, Raymond J.C.  
; APPLICANT: Bagley, Angela L.  
; TITLE OF INVENTION: No. 5262158el Bacillus thuringiensis Isolates for  
; NUMBER OF SEQUENCES: 30  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: David R. Saliwanchik  
; STREET: 2421 N.W. 41st Street, Suite A-1  
; CITY: Gainesville  
; STATE: FL  
; COUNTRY: USA  
; ZIP: 32606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/876,280  
; FILING DATE: 19920430  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Saliwanchik, David R.  
; REGISTRATION NUMBER: 31,794  
; REFERENCE/DOCKET NUMBER: M/S 104  
; TELEPHONE: 904-375-8100  
; TELEFAX: 904-372-5800

; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 395 amino acids  
; TYPE: AMINO ACID  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: YES  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: BACILLUS THURINGIENSIS  
; INDIVIDUAL ISOLATE: PS69D1  
; IMMEDIATE SOURCE:  
; CLONE: E. coli NM522(pMYC2317) NRRL B-18816  
; FEATURE:  
; NAME/KEY: Protein  
; LOCATION: 1..395  
; US-07-876-280-9

Query Match 31.8%; Score 41; DB 1; Length 395;  
Best Local Similarity 40.9%; Pred. No. 1.9e+02;  
Matches 9; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

QY 5 EALADOTDALQSEEAAYVKADN 26  
DB 231 EELLKKVDDLKKELEAAIKAE 252

RESULT 40  
US-08-049-783-6  
; Sequence 6, Application US/08049783  
; Patent No. 5439881  
; GENERAL INFORMATION:  
; APPLICANT: Narva, Kenneth E  
; APPLICANT: Schwab, George E  
; APPLICANT: Payne, Jewel M  
; TITLE OF INVENTION: Gene Encoding No. 5439881el Nematode-Active  
; NUMBER OF SEQUENCES: 18  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Jeff Lloyd  
; STREET: 2421 N.W. 41st Street  
; CITY: Gainesville  
; STATE: FL  
; COUNTRY: USA  
; ZIP: 32606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/049,783  
; FILING DATE: 19930419  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Lloyd, Jeff  
; TELEPHONE: 904-375-8100  
; TELEFAX: 904-372-5800  
; INFORMATION FOR SEQ ID NO: 6 (PS69D1):  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 395 amino acids  
; TYPE: AMINO ACID  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: YES  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: BACILLUS THURINGIENSIS  
; INDIVIDUAL ISOLATE: PS69D1  
; IMMEDIATE SOURCE:

QY	2	DALEALADQTDALQSEEA	AAVVKADNAA	28
		:  :	:      :  :	
Db	3434	DALRMAEAGAEVQIVEAD	VARRDDVA	3460

```

RESULT 35
US-09-568-486-6
; Sequence 6, Application US/09568486
; Patent No. 6355459
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,486
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 3798
; TYPE: prt
; ORGANISM: Sorangium cellulosum
US-09-568-486-6

```

```

RESULT 36
US-09-568-472-6
; Sequence 6, Application US/09568472
; Patent No. 6358719
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,472
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 3798
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-568-472-6

```

RESULT 37  
 US-08-669-408B-8  
 : Sequence 8, Application US/08669408B  
 : Patent No. 610055  
 : GENERAL INFORMATION:  
 : APPLICANT: GUSS, Bengt  
 : APPLICANT: JONSSON, Hans  
 : APPLICANT: LINDBERG, Martin  
 : APPLICANT: MUELLER, Hans-Peter  
 : APPLICANT: RANTAMAKI, Liisa K.  
 : TITLE OF INVENTION: METHOD AND MEANS FOR PRODUCING  
 : TITLE OF INVENTION: PLASMAPROTEINASE INHIBITOR-BINDING PROTEINS  
 : NUMBER OF SEQUENCES: 12  
 : CORRESPONDENCE ADDRESS:  
 : ADDRESSEE: Foley & Lardner  
 : STREET: 3000 K Street, N.W., Suite 500  
 : CITY: Washington  
 : STATE: D.C.  
 : COUNTRY: USA  
 : ZIP: 20007-5109  
 : COMPUTER READABLE FORM:  
 : MEDIUM TYPE: Floppy disk  
 : COMPUTER: IBM PC compatible  
 : OPERATING SYSTEM: PC-DOS/MS-DOS  
 : SOFTWARE: PatentIn Release #1.0, Version #1.30  
 : CURRENT APPLICATION DATA:  
 : APPLICATION NUMBER: US/08/669,408B  
 : FILING DATE: 03-JUL-1996  
 : CLASSIFICATION:  
 : PRIOR APPLICATION DATA:  
 : APPLICATION NUMBER: WO PCT/SE94/00826  
 : FILING DATE: 06-SEP-1994  
 : PRIOR APPLICATION DATA:  
 : APPLICATION NUMBER: SE 9302855-3  
 : FILING DATE: 06-SEP-1993  
 : ATTORNEY/AGENT INFORMATION:  
 : NAME: BENT, Stephen A.  
 : REGISTRATION NUMBER: 29,768  
 : REFERENCE/DOCKET NUMBER: 61743/102  
 : TELECOMMUNICATION INFORMATION:  
 : TELEPHONE: (202)672-5300  
 : TELEFAX: (202)672-5399  
 : TELEX: 904136  
 : INFORMATION FOR SEQ ID NO: 8:  
 : SEQUENCE CHARACTERISTICS:  
 : LENGTH: 173 amino acids  
 : TYPE: amino acid  
 : TOPOLOGY: linear  
 : MOLECULE TYPE: protein  
 : US-08-669-408B-8

```

RESULT 38
PCT-US93-01676A-4
; Sequence 4, Application PC/TUS9301676A
; GENERAL INFORMATION:
; APPLICANT: Stambrook, Peter J.
; TITLE OF INVENTION: MUTAGENICITY TESTING USING REPORTER
; TITLE OF INVENTION: GENES WITH MODIFIED METHYLATION FREQUENCIES
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Browdy and Neimark
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.

```



QY 2 DALEALADOTDALQSEEAAYVKAD 25  
||||| || | : |||  
Db 1356 DALEAL---DRALDRETTVVAD 1376

## RESULT 27

US-08-710-749-8  
; Sequence 8, Application US/08710749  
; Patent No. 5955089  
; GENERAL INFORMATION:  
; APPLICANT: Briles, David E.  
; APPLICANT: Hollingshead, Susan  
; APPLICANT: Becker, Robert  
; TITLE OF INVENTION: STRAIN SELECTION OF PNEUMOCOCCAL SURFACE  
; TITLE OF INVENTION: PROTEINS  
; NUMBER OF SEQUENCES: 28  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Curtis, Morris & Safford  
; STREET: 530 Fifth Avenue  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10036

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/710,749  
; FILING DATE: 20-SEP-1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Frommer, William S.  
; REGISTRATION NUMBER: 25,506  
; REFERENCE/DOCKET NUMBER: 454312-2074  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 840-3333  
; TELEFAX: (212) 840-0712  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 102 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: n/a  
; TOPOLOGY: linear  
; MOLECULE TYPE: amino acid  
US-08-710-749-8

Query Match 32.6%; Score 42; DB 2; Length 102;  
Best Local Similarity 45.5%; Pred. No. 27;  
Matches 10; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY 4 LEALADOTDALQSEEAAYVKAD 25  
|| | : ||| : |||  
Db 36 LEELSKIDELDAEIAKLNKDD 57

## RESULT 28

US-09-521-668B-18  
; Sequence 18, Application US/09521668B  
; Patent No. 6303383  
; GENERAL INFORMATION:  
; APPLICANT: NAKAMURA, JUN  
; APPLICANT: KANNO, SOHEI  
; APPLICANT: KIMURA, EIICHIRO  
; APPLICANT: MATSUI, KAUIKO  
; APPLICANT: NAKAMATSU, TSUYOSHI  
; TITLE OF INVENTION: TEMPERATURE SENSITIVE PLASMID FOR CORYNEFORM BACTERIA  
; FILE REFERENCE: 0010-1093-0  
; CURRENT APPLICATION NUMBER: US/09/521,668B  
; CURRENT FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: JP 11-69896  
; PRIOR FILING DATE: 1999-03-16

; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 18  
; LENGTH: 427  
; TYPE: PRT  
; ORGANISM: Brevibacterium lactofermentum  
US-09-521-668B-18

Query Match 32.6%; Score 42; DB 4; Length 427;  
Best Local Similarity 47.4%; Pred. No. 1.5e+02;  
Matches 9; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 4 LEALADOTDALQSEEAAYV 22  
| : ||||| : : | : |  
Db 256 LDMLAQSDAGEDMDAVLV 274

## RESULT 29

US-09-521-668B-20  
; Sequence 20, Application US/09521668B  
; Patent No. 6303383  
; GENERAL INFORMATION:  
; APPLICANT: NAKAMURA, JUN  
; APPLICANT: KANNO, SOHEI  
; APPLICANT: KIMURA, EIICHIRO  
; APPLICANT: MATSUI, KAUIKO  
; APPLICANT: NAKAMATSU, TSUYOSHI  
; TITLE OF INVENTION: TEMPERATURE SENSITIVE PLASMID FOR CORYNEFORM BACTERIA  
; FILE REFERENCE: 0010-1093-0  
; CURRENT APPLICATION NUMBER: US/09/521,668B  
; CURRENT FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: JP 11-69896  
; PRIOR FILING DATE: 1999-03-16  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 20  
; LENGTH: 427  
; TYPE: PRT  
; ORGANISM: Brevibacterium lactofermentum  
US-09-521-668B-20

Query Match 32.6%; Score 42; DB 4; Length 427;  
Best Local Similarity 47.4%; Pred. No. 1.5e+02;  
Matches 9; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 4 LEALADOTDALQSEEAAYV 22  
| : ||||| : : | : |  
Db 256 LDMLAQSDAGEDMDAVLV 274

## RESULT 30

US-09-413-814-67  
; Sequence 67, Application US/09413814  
; Patent No. 6225064  
; GENERAL INFORMATION:  
; APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH  
; APPLICANT: Bristol-Myers Squibb, Co.  
; APPLICANT: Beyer, Stefan  
; APPLICANT: Bioecker, Helmut  
; APPLICANT: Brandt, Petra  
; APPLICANT: Cino, Paul M  
; APPLICANT: Dougherty, Brian A  
; APPLICANT: Goldberg, Steven L  
; APPLICANT: Hofle, Gerhard  
; APPLICANT: Mueller, Joachim  
; APPLICANT: Reichenbach, Hans  
; TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polyketide or  
; FILE REFERENCE: PCT/US 99/23535  
; CURRENT APPLICATION NUMBER: US/09/413,814  
; CURRENT FILING DATE: 1999-10-07  
; EARLIER APPLICATION NUMBER: DE 198 46 493.2  
; EARLIER FILING DATE: 1998-10-09

```
; STATE: PA
; COUNTRY: US
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/238.557
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/934,846
; FILING DATE:
; NAME: Dickinson, Todd O
; REGISTRATION NUMBER: 28,354
; REFERENCE/DOCKET NUMBER: GM10088
; TELEPHONE: 215-994-2252
; TELEFAX: 215-994-2222
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 418 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-238-557-2

Query Match 33.3%; Score 43; DB 4; Length 418;
Best Local Similarity 33.3%; Pred. No. 1e+02;
Matches 9; Conservative 5; Mismatches 13; Indels 0; Gaps 0;

QY 2 DALEALADQTDALQSEEAAYVKADNA 28
| | | | | : : : : |
Db 167 DHQETLGSLEAIAPQKAGIPKAGKA 193

RESULT 25
US-08-764-100-25
; Sequence 25, Application US/08764100
; Patent No. 5773700
; GENERAL INFORMATION:
; APPLICANT: van Grinsven J., Martinus Q.
; APPLICANT: De Haan, Petrus T.
; APPLICANT: Gielen L., Johannes J.
; APPLICANT: Peters, Dirk
; APPLICANT: Goldbach, Robert W.
; TITLE OF INVENTION: Improvements in or Relating to Organic
; TITLE OF INVENTION: Compounds
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sandoz Agro, Inc
; STREET: 975 California Avenue
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/764.100
; FILING DATE: 06-DEC-1996
; CLASSIFICATION: 800
; PRIOR APPLICATION NUMBER: US/08/214.064
; FILING DATE:
; APPLICATION NUMBER: US 08/032.235
```

```
; FILING DATE: 17-MAR-1993
; APPLICATION NUMBER: GB 9206016.9
; FILING DATE: 19-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5773700ris, Allen E.
; REGISTRATION NUMBER: 34,490
; REFERENCE/DOCKET NUMBER: 137-1061
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 354-3592
; TELEFAX: (415) 857-1125
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 744 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-764-100-25

Query Match 33.3%; Score 43; DB 1; Length 744;
Best Local Similarity 38.5%; Pred. No. 2.1e+02;
Matches 10; Conservative 6; Mismatches 10; Indels 0; Gaps 0;

QY 1 SDALEALADQTDALQSEEAAYVKADN 26
| : | | : : | | | : : |
Db 432 SERLEPRLEILEALASERVALLEASN 457

RESULT 26
US-08-804-227C-5
; Sequence 5, Application US/08804227C
; Patent No. 5876991
; GENERAL INFORMATION:
; APPLICANT: DeHoff, Bradley S.
; APPLICANT: Kuhstoss, Stuart A.
; APPLICANT: Rostock, Paul R., Jr.
; APPLICANT: Sutton, Kimberly L.
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: THOMAS G. PLANT 1501
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCII(DOS) Text only
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804.227C
; FILING DATE: February 21, 1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Plant, Thomas, G.
; REGISTRATION NUMBER: 35,784
; REFERENCE/DOCKET NUMBER: X-8231
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-2459
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1611 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-804-227C-5

Query Match 32.9%; Score 42.5; DB 2; Length 1611;
Best Local Similarity 54.2%; Pred. No. 6.1e+02;
Matches 13; Conservative 1; Mismatches 7; Indels 3; Gaps 1;
```

/ APPLICATION NUMBER:  
/ FILING DATE:  
/ ATTORNEY/AGENT INFORMATION:  
/ NAME: Dickinson, Todd Q  
/ REGISTRATION NUMBER: 28,354  
/ REFERENCE/DOCKET NUMBER: GM10088  
/ TELEPHONE: 215-994-2252  
/ TELEFAX: 215-994-2222  
/ TELEX:  
/ INFORMATION FOR SEQ ID NO: 4:  
/ SEQUENCE CHARACTERISTICS:  
/ LENGTH: 407 amino acids  
/ TYPE: amino acid  
/ STRANDEDNESS: single  
/ TOPOLOGY: linear  
/ US-08-934-846-4

Query Match 33.3%; Score 43; DB 2; Length 407;  
Best Local Similarity 33.3%; Pred. No. 1e+02;  
Matches 9; Conservative 5; Mismatches 13; Indels 0; Gaps 0;

QY 2 DALEALADQTDALQSEEAHVVKADNAA 28  
| | | | | : | : | : | | |  
DB 167 DHOETLGDSLEATAEQKAGIFKAGKKA 193

RESULT 22  
US-09-238-557-4  
/ Sequence 4, Application US/09238557  
/ Patent No. 6165472  
/ GENERAL INFORMATION:  
/ APPLICANT: Pearson, Stewart C.  
/ APPLICANT: Greenwood, Rebecca C.  
/ TITLE OF INVENTION: NOVEL FOIC  
/ NUMBER OF SEQUENCES: 4  
/ CORRESPONDENCE ADDRESS:  
/ ADDRESSEE: Dechert Price & Rhoads  
/ STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre  
/ CITY: Philadelphia  
/ STATE: PA  
/ COUNTRY: US  
/ ZIP: 19103  
/ COMPUTER READABLE FORM:  
/ MEDIUM TYPE: Diskette  
/ COMPUTER: IBM Compatible  
/ OPERATING SYSTEM: DOS  
/ SOFTWARE: FastSeq for Windows Version 2.0  
/ CURRENT APPLICATION DATA:  
/ APPLICATION NUMBER: US/09/238,557  
/ FILING DATE:  
/ CLASSIFICATION:  
/ PRIOR APPLICATION DATA:  
/ APPLICATION NUMBER: 08/934,846  
/ FILING DATE:  
/ ATTORNEY/AGENT INFORMATION:  
/ NAME: Dickinson, Todd Q  
/ REGISTRATION NUMBER: 28,354  
/ REFERENCE/DOCKET NUMBER: GM10088  
/ TELEPHONE: 215-994-2252  
/ TELEFAX: 215-994-2222  
/ TELEX:  
/ INFORMATION FOR SEQ ID NO: 4:  
/ SEQUENCE CHARACTERISTICS:  
/ LENGTH: 407 amino acids  
/ TYPE: amino acid  
/ STRANDEDNESS: single  
/ TOPOLOGY: linear  
/ US-09-238-557-4

Query Match 33.3%; Score 43; DB 4; Length 407;  
Best Local Similarity 33.3%; Pred. No. 1e+02;

Matches 9; Conservative 5; Mismatches 13; Indels 0; Gaps 0;  
QY 2 DALEALADQTDALQSEEAHVVKADNAA 28  
| | | | | : | : | : | | |  
DB 167 DHOETLGDSLEATAEQKAGIFKAGKKA 193

RESULT 23  
US-08-934-846-2  
/ Sequence 2, Application US/08934846  
/ Patent No. 5882898  
/ GENERAL INFORMATION:  
/ APPLICANT: Pearson, Stewart C.  
/ APPLICANT: Greenwood, Rebecca C.  
/ TITLE OF INVENTION: NOVEL FOIC  
/ NUMBER OF SEQUENCES: 4  
/ CORRESPONDENCE ADDRESS:  
/ ADDRESSEE: Dechert Price & Rhoads  
/ STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre  
/ CITY: Philadelphia  
/ STATE: PA  
/ COUNTRY: US  
/ ZIP: 19103  
/ COMPUTER READABLE FORM:  
/ MEDIUM TYPE: Diskette  
/ COMPUTER: IBM Compatible  
/ OPERATING SYSTEM: DOS  
/ SOFTWARE: FastSeq for Windows Version 2.0  
/ CURRENT APPLICATION DATA:  
/ APPLICATION NUMBER: US/08/934,846  
/ FILING DATE:  
/ CLASSIFICATION: 536  
/ PRIOR APPLICATION DATA:  
/ APPLICATION NUMBER:  
/ FILING DATE:  
/ ATTORNEY/AGENT INFORMATION:  
/ NAME: Dickinson, Todd Q  
/ REGISTRATION NUMBER: 28,354  
/ REFERENCE/DOCKET NUMBER: GM10088  
/ TELEPHONE: 215-994-2252  
/ TELEFAX: 215-994-2222  
/ TELEX:  
/ INFORMATION FOR SEQ ID NO: 2:  
/ SEQUENCE CHARACTERISTICS:  
/ LENGTH: 418 amino acids  
/ TYPE: amino acid  
/ STRANDEDNESS: single  
/ TOPOLOGY: linear  
/ US-08-934-846-2

Query Match 33.3%; Score 43; DB 2; Length 418;  
Best Local Similarity 33.3%; Pred. No. 1e+02;  
Matches 9; Conservative 5; Mismatches 13; Indels 0; Gaps 0;

QY 2 DALEALADQTDALQSEEAHVVKADNAA 28  
| | | | | : | : | : | | |  
DB 167 DHOETLGDSLEATAEQKAGIFKAGKKA 193

RESULT 24  
US-09-238-557-2  
/ Sequence 2, Application US/09238557  
/ Patent No. 6165472  
/ GENERAL INFORMATION:  
/ APPLICANT: Pearson, Stewart C.  
/ APPLICANT: Greenwood, Rebecca C.  
/ TITLE OF INVENTION: NOVEL FOIC  
/ NUMBER OF SEQUENCES: 4  
/ CORRESPONDENCE ADDRESS:  
/ ADDRESSEE: Dechert Price & Rhoads  
/ STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre  
/ CITY: Philadelphia

```
RESULT 19
US-08-235-838-7
; Sequence 7, Application US/08235838
; Patent No. 5571894
; GENERAL INFORMATION:
; APPLICANT: Wels, Winfried S.
; APPLICANT: Hynes, Nancy E.
; APPLICANT: Harwerth, Ina-Maria
; APPLICANT: Groner, Bernd
; APPLICANT: Hardman, No. 5571894man
; APPLICANT: Zwickl, Markus
; TITLE OF INVENTION: Recombinant Antibodies Specific for a
; TITLE OF INVENTION: Growth Factor Receptor
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: New York
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/235,838
; FILING DATE: TBA
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/828,832
; FILING DATE: 31-JAN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 91-810079.3
; FILING DATE: 05-FEB-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; REFERENCE/DOCKET NUMBER: 4-18518/A/CIP/CONT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (919)541-8614
; TELEFAX: (919)541-8689
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 711 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-235-838-7

Query Match 33.7%; Score 43.5; DB 1; Length 711;
Best Local Similarity 31.7%; Pred. No. 1.6e+02;
Matches 13; Conservative 6; Mismatches 9; Indels 13; Gaps 1;

QY 1 SDALEALADOTDA-----LQSEEAAVVKADNAA 28
:|::|||
Db 555 NDSVPTLAQMTDKAIELLSKNEKGFFLQVEGASIDKQDHAA 595

RESULT 20
US-08-465-473B-7
; Sequence 7, Application US/08465473B
; Patent No. 5939531
; GENERAL INFORMATION:
; APPLICANT: Wels, Winfried S.
; APPLICANT: Hynes, Nancy E.
; APPLICANT: Harwerth, Ina-Maria
; APPLICANT: Groner, Bernd
; APPLICANT: Hardman, No. 5939531man
; APPLICANT: Zwickl, Markus
; TITLE OF INVENTION: Recombinant Antibodies Specific for a
; TITLE OF INVENTION: Growth Factor Receptor
```

```
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NOVARTIS Corporation
; STREET: 564 Morris Avenue
; CITY: Summit
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07901-6940
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/465,473B
; FILING DATE: 5 June 1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/828,832
; FILING DATE: 31-JAN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 91-810079.3
; FILING DATE: 05-FEB-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Pfeiffer, Hesna J.
; REGISTRATION NUMBER: 22,640
; REFERENCE/DOCKET NUMBER: 4-18518/A/CIP/CONT2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (908)522 6940
; TELEFAX: (908)522 6955
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 711 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-465-473B-7

Query Match 33.7%; Score 43.5; DB 2; Length 711;
Best Local Similarity 31.7%; Pred. No. 1.6e+02;
Matches 13; Conservative 6; Mismatches 9; Indels 13; Gaps 1;

QY 1 SDALEALADOTDA-----LQSEEAAVVKADNAA 28
:|::|||
Db 555 NDSVPTLAQMTDKAIELLSKNEKGFFLQVEGASIDKQDHAA 595

RESULT 21
US-08-934-846-4
; Sequence 4, Application US/08934846
; Patent No. 5882898
; GENERAL INFORMATION:
; APPLICANT: Pearson, Stewart C.
; APPLICANT: Greenwood, Rebecca C.
; TITLE OF INVENTION: NOVEL FOLC
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
; COUNTRY: US
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/934,846
; FILING DATE:
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
```

CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/100,708  
FILING DATE: July 29, 1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Wong, Wean Khing  
REGISTRATION NUMBER: 33,561  
REFERENCE/DOCKET NUMBER: 5324.US.P1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (708) 938-3517  
TELEFAX: (708) 938-2623  
TELEX:  
INFORMATION FOR SEQ ID NO: 31:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 471 amino acid residues  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
ORIGINAL SOURCE:  
ORGANISM:  
US-08-657-392-31

Query Match 33.7% Score 43.5; DB 2; Length 471;  
Best Local Similarity 31.7% Pred. No. 1e+02;  
Matches 13; Conservative 6; Mismatches 6; Indels 13; Gaps 1;

Oy 1 SDALEALADQTD-----LQSEEAHVVKADNAA 28  
Db 315 NDSVPTLAQMTDKATELLSKNEKGFLLQVEGASIDKQDHA 355

## RESULT 17

PCT-US94-02539-31  
Sequence 31, Application PC/TUS9402539  
GENERAL INFORMATION:  
APPLICANT: Brate, E.M.  
APPLICANT: Brennan, C.A.  
APPLICANT: Bridon, D.P.  
APPLICANT: Jaffe, K.D.  
APPLICANT: Krafft, G.A.  
APPLICANT: Mandeckl, W.  
APPLICANT: March, S.C.  
APPLICANT: Russell, J.R.  
APPLICANT: Yue, V.T.  
TITLE OF INVENTION: Genetically Engineered Enzymes  
TITLE OF INVENTION: And Their  
NUMBER OF INVENTION: Conjugates For Diagnostic Assays  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: ABBOTT LABORATORIES  
STREET: One Abbott Park Road  
CITY: Abbott Park  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60064-3500  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: SoftPC  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US94/02539  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Wong, Wean Khing  
REGISTRATION NUMBER: 33,561  
REFERENCE/DOCKET NUMBER: 5324.PC.01  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (708) 938-3517  
TELEFAX: (708) 938-2623  
TELEX:

INFORMATION FOR SEQ ID NO: 31:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 471 amino acid residues  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
ORIGINAL SOURCE:  
ORGANISM:  
PCT-US94-02539-31

Query Match 33.7% Score 43.5; DB 5; Length 471;  
Best Local Similarity 31.7% Pred. No. 1e+02;  
Matches 13; Conservative 6; Mismatches 13; Indels 13; Gaps 1;

Oy 1 SDALEALADQTD-----LQSEEAHVVKADNAA 28  
Db 315 NDSVPTLAQMTDKATELLSKNEKGFLLQVEGASIDKQDHA 355

## RESULT 18

US-08-928-213B-8  
Sequence 8, Application US/08928213B  
Patent No. 6238905  
GENERAL INFORMATION:  
APPLICANT: McHenry, Charles S.  
Seville, Mark  
Cull, Millard G.  
TITLE OF INVENTION: NOVEL THERMOPHILIC POLYMERASE III  
HOLOENZYME  
NUMBER OF SEQUENCES: 195  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MEDLEN & CARROLL, LLP  
STREET: 220 Montgomery Street, Suite 2200  
CITY: San Francisco  
STATE: CA  
COUNTRY: USA  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/928,213B  
FILING DATE: 12-Sep-1997  
CLASSIFICATION: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: MacKnight, Kamrin T.  
REGISTRATION NUMBER: 38,230  
REFERENCE/DOCKET NUMBER: ENZYCO-02550  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-705-8410  
TELEFAX: 415-397-8338  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 528 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: not relevant  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 8:  
US-08-928-213B-8

Query Match 33.7% Score 43.5; DB 4; Length 528;  
Best Local Similarity 46.2% Pred. No. 1.2e+02;  
Matches 12; Conservative 7; Mismatches 6; Indels 1; Gaps 1;

Oy 2 DALEALADQTDALQSEEAHVVKADNA 27  
Db 314 EAMERLARRSDAL-SLEVALLEAGRA 338



```
?
? APPLICANT: March, S.C.
? APPLICANT: Russell, J.R.
? APPLICANT: Yue, V.T.
? TITLE OF INVENTION: Genetically Engineered Enzymes And Their
? CONJUGATES FOR DIAGNOSTIC ASSAYS
? NUMBER OF SEQUENCES: 34
? CORRESPONDENCE ADDRESS:
? STREET: One Abbott Park Road
? CITY: Abbott Park
? STATE: Illinois
? COUNTRY: USA
? ZIP: 60064-3500
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: SoftPC
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/657,392
? FILING DATE:
? CLASSIFICATION: 435
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: 08/100,708
? FILING DATE: July 29, 1993
? ATTORNEY/AGENT INFORMATION:
? NAME: Wong, Wean Khing
? REGISTRATION NUMBER: 33,561
? REFERENCE/DOCKET NUMBER: 5324.US.PI
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (708) 938-3517
? TELEFAX: (708) 938-2623
? TELEX:
? INFORMATION FOR SEQ ID NO: 2:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 449 amino acid residues
? TYPE: amino acid
? STRANDEDNESS:
? TOPOLOGY: unknown
? MOLECULE TYPE: protein
? ORIGINAL SOURCE:
? ORGANISM:
? PCT-US94-02539-2
?
? Query Match 33.7%; Score 43.5; DB 2; Length 449;
? Best Local Similarity 31.7%; Pred. No. 95;
? Matches 13; Conservative 6; Mismatches 9; Indels 13; Gaps 1;
?
? QY 1 SDALEALADQTD-----LQSEEAAYVKADNAA 28
? Db 293 NDSVPTLAQMTDKAIELLSKNEKGFFLQVBGASIDKQDHAA 333
?
? RESULT 15
? PCT-US94-02539-2
? GENERAL INFORMATION:
? APPLICANT: Brate, E.M.
? APPLICANT: Brennan, C.A.
? APPLICANT: Bridon, D.P.
? APPLICANT: Jaffe, K.D.
? APPLICANT: Krafft, G.A.
? APPLICANT: Mandeck, W.
? APPLICANT: March, S.C.
? APPLICANT: Russell, J.R.
? APPLICANT: Yue, V.T.
? TITLE OF INVENTION: Genetically Engineered Enzymes
? AND THEIR CONJUGATES FOR DIAGNOSTIC ASSAYS
? NUMBER OF SEQUENCES: 34
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: ABBOTT LABORATORIES
? STREET: One Abbott Park Road
? CITY: Abbott Park
? STATE: Illinois
? COUNTRY: USA
? ZIP: 60064-3500
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: SoftPC
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/657,392
? FILING DATE:
```

```
?
? CITY: Abbott Park
? STATE: Illinois
? COUNTRY: USA
? ZIP: 60064-3500
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: SoftPC
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: PCT/US94/02539
? FILING DATE:
? CLASSIFICATION:
? ATTORNEY/AGENT INFORMATION:
? NAME: Wong, Wean Khing
? REGISTRATION NUMBER: 33,561
? REFERENCE/DOCKET NUMBER: 5324.PC.01
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (708) 938-3517
? TELEFAX: (708) 938-2623
? TELEX:
? INFORMATION FOR SEQ ID NO: 2:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 449 amino acid residues
? TYPE: amino acid
? STRANDEDNESS:
? TOPOLOGY: unknown
? MOLECULE TYPE: protein
? ORIGINAL SOURCE:
? ORGANISM:
? PCT-US94-02539-2
?
? Query Match 33.7%; Score 43.5; DB 5; Length 449;
? Best Local Similarity 31.7%; Pred. No. 95;
? Matches 13; Conservative 6; Mismatches 9; Indels 13; Gaps 1;
?
? QY 1 SDALEALADQTD-----LQSEEAAYVKADNAA 28
? Db 293 NDSVPTLAQMTDKAIELLSKNEKGFFLQVBGASIDKQDHAA 333
?
? RESULT 16
? US-08-657-392-31
? Sequence 31, Application US/08657392
? Patent No. 5843634
? GENERAL INFORMATION:
? APPLICANT: Brate, E.M.
? APPLICANT: Brennan, C.A.
? APPLICANT: Bridon, D.P.
? APPLICANT: Jaffe, K.D.
? APPLICANT: Krafft, G.A.
? APPLICANT: Mandeck, W.
? APPLICANT: March, S.C.
? APPLICANT: Russell, J.R.
? APPLICANT: Yue, V.T.
? TITLE OF INVENTION: Genetically Engineered Enzymes And Their
? CONJUGATES FOR DIAGNOSTIC ASSAYS
? NUMBER OF SEQUENCES: 34
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: ABBOTT LABORATORIES
? STREET: One Abbott Park Road
? CITY: Abbott Park
? STATE: Illinois
? COUNTRY: USA
? ZIP: 60064-3500
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: SoftPC
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/657,392
? FILING DATE:
```

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US94/04173  
FILING DATE: 15-APR-1994  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/049,282  
FILING DATE: 16-APR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Bastian, Kevin L.  
REGISTRATION NUMBER: 34,774  
REFERENCE/DOCKET NUMBER: 12176-34-1PC  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 543-9600  
TELEFAX: (415) 543-5043  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 382 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US94-04173-2

Query Match 33.7% Score 43.5; DB 5; Length 382;  
Best Local Similarity 52.4% Pred. No. 78;  
Matches 11; Conservative 5; Mismatches 4; Indels 1; Gaps 1;

OY 4 LEALADQTDALQSEEAAYVK 23  
||| I:|:| |::: ||  
Db 243 LEARNQTEATLQTKDTGTVK 263

## RESULT 12

US-08-537-715-4  
Sequence 4, Application US/08537715  
Patent No. 5910627  
GENERAL INFORMATION:  
APPLICANT: Chuck, George S.  
APPLICANT: Dooner, Hugo K.  
APPLICANT: Courtney-Guterson, Neal  
APPLICANT: Keller, Janis  
APPLICANT: Nijjar, Charanjit S.  
APPLICANT: Ralston, Edward J.  
TITLE OF INVENTION: PH Genes and Their Uses  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, Eighth Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/537,715  
FILING DATE: 16-FEB-1996  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/049,282  
FILING DATE: 16-APR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/US94/04173  
FILING DATE: 15-APR-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Bastian, Kevin L.  
REGISTRATION NUMBER: 34,774  
REFERENCE/DOCKET NUMBER: 012176-003410US  
TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 383 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-537-715-4

Query Match 33.7% Score 43.5; DB 2; Length 383;  
Best Local Similarity 52.4% Pred. No. 78;  
Matches 11; Conservative 5; Mismatches 4; Indels 1; Gaps 1;

OY 4 LEALADQTDALQSEEAAYVK 23  
||| I:|:| |::: ||  
Db 244 LEARNQTEATLQTKDTGTVK 264

## RESULT 13

PCT-US94-04173-4  
Sequence 4, Application PC/TUS9404173  
GENERAL INFORMATION:  
APPLICANT:  
TITLE OF INVENTION: PH GENES AND THEIR USES  
NUMBER OF SEQUENCES: 4  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US94/04173  
FILING DATE: 15-APR-1994  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/049,282  
FILING DATE: 16-APR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Bastian, Kevin L.  
REGISTRATION NUMBER: 34,774  
REFERENCE/DOCKET NUMBER: 12176-34-1PC  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 543-9600  
TELEFAX: (415) 543-5043  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 383 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US94-04173-4

Query Match 33.7% Score 43.5; DB 5; Length 383;  
Best Local Similarity 52.4% Pred. No. 78;  
Matches 11; Conservative 5; Mismatches 4; Indels 1; Gaps 1;

OY 4 LEALADQTDALQSEEAAYVK 23  
||| I:|:~| |::: ||  
Db 244 LEARNQTEATLQTKDTGTVK 264

## RESULT 14

US-08-657-392-2  
Sequence 2, Application US/08657392  
Patent No. 5843634  
GENERAL INFORMATION:  
APPLICANT: Brate, E.M.  
APPLICANT: Brennan, C.A.  
APPLICANT: Bridon, D.P.  
APPLICANT: Jaffe, K.D.  
APPLICANT: Kraft, G.A.  
APPLICANT: Mandeckl, W.

```
?  
? APPLICATION: PH GENES AND THEIR USES  
?  
? TITLE OF INVENTION: 4  
?  
? NUMBER OF SEQUENCES: 4  
?  
? COMPUTER READABLE FORM:  
? MEDIUM TYPE: Floppy disk  
? COMPUTER: IBM PC compatible  
?
```

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/320,774
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/762,106
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brotman, Harris F.
; REGISTRATION NUMBER: 35,461
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 654-2428
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 527 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: Internal
; US-09-320-774-9

```

Query Match	34.5%	Score 44.5;	DB 4;	Length 527;
Best Local Similarity	51.7%	Pred. No. 82;		
Matches 15;	Conservative	4;	Mismatches	7;
			Indels	3;
			Gaps	2;

QY 1 SDALEALADQTDLAQSEEAUVK--ADNA 27  
          :||||| : -||:||||| -|||  
Db 380 ADAVEAAAAELTA-QAEAAANAKWEADKA 407

```

RESULT 6
US-08-745-404-3
; Sequence 3, Application US/08745404B
; Patent No. 6096717
; GENERAL INFORMATION:
; APPLICANT: Jarvik, Jonathan W.
; TITLE OF INVENTION: Method For Producing Tagged Genes,,
; TITLE OF INVENTION: Transcripts And Proteins
; FILE REFERENCE: 2087-961422
; CURRENT APPLICATION NUMBER: US/08/745,404B
; CURRENT FILING DATE: 1996-11-08
; EARLIER APPLICATION NUMBER: 08/000,619
; EARLIER FILING DATE: 1993-01-05
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 552
; TYPE: PRT
; ORGANISM: Chlamydomonas
US-08-745-404-3

```

Query Match	34.5%	Score 44.5;	DB 3;	Length 552;
Best Local Similarity	51.7%	Pred. No. 87;		
Matches 15;	Conservative	4;	Mismatches 7;	Indels 3;
				Gaps 2;

QY 1 SDALEALADQTDALQSEEAUVK--ADNA 27  
:|||||:|:|||||:|||||

DB 405 ADAVEAAAAELTA-QAEEAANAKWEADKA 432

```

RESULT 7
; US-08-961-083-192
; Sequence 192, Application us/08961083
; Patent No. 6159469
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
;

```

```

? CITY: Rockville
? STATE: Maryland
? COUNTRY: USA
? ZIP: 20850
?
? COMPUTER READABLE FORM:
?
? MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
?
? COMPUTER: HP Vectra 486/33
? OPERATING SYSTEM: MSDOS version 6.2
? SOFTWARE: ASCII Text
?
? CURRENT APPLICATION DATA:
?
? APPLICATION NUMBER: US/08/961,083
?
? FILING DATE:
?
? CLASSIFICATION: 435
?
? PRIOR APPLICATION DATA:
?
? APPLICATION NUMBER:
?
? FILING DATE:
?
? ATTORNEY/AGENT INFORMATION:
?
? NAME: Brookes, A. Anders
?
? REGISTRATION NUMBER: 36,373
?
? REFERENCE/DOCKET NUMBER: PB340P2
?
? TELECOMMUNICATION INFORMATION:
?
? TELEPHONE: (301) 309-8504
?
? TELEFAX: (301) 309-8512
?
? INFORMATION FOR SEQ ID NO: 192:
?
? SEQUENCE CHARACTERISTICS:
?
? LENGTH: 344 amino acids
?
? TYPE: amino acid
?
? STRANDEDNESS: single
?
? TOPOLOGY: linear
?
? MOLECULE TYPE: protein
?
? US-08-961-083-192

```

```
Query Match      34.1%; Score 44; DB 4; Length 344;
Best Local Similarity 55.0%; Pred. No. 58;
Matches 11; Conservative 2; Mismatches 7; Indels
```

---

```
QY      8 ADQTDLAQSEEAAYVKADNA 27
          | | | | | | | | | | | |
Db       3 AGOTDASOIEKAAYSOGGKA 22
```

```

RESULT 8
US-08-960-022-14
; Sequence 14, Application US/08960022
; Patent No. 5976837
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M.
; APPLICANT: LaValle, Edward R.
; APPLICANT: Racle, Lisa A.
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Spaulding, Vikki
; APPLICANT: Agostino, Michael J.
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
; TITLE OF INVENTION: ENCODING THEM
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: MA
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/960,022
; FILING DATE:
; CLASSIFICATION: 514

```

; Sequence 2, Application US/08745404B  
; Patent No. 6096717  
; GENERAL INFORMATION:  
; APPLICANT: Jarvik, Jonathan W.  
; TITLE OF INVENTION: Method For Producing Tagged Genes,  
; FILE OF INVENTION: Transcripts And Proteins  
; FILE REFERENCE: 2087-961422  
; CURRENT APPLICATION NUMBER: US/08/745.404B  
; CURRENT FILING DATE: 1996-11-08  
; EARLIER APPLICATION NUMBER: 08/000.619  
; EARLIER FILING DATE: 1993-01-05  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 516  
; TYPE: PRT  
; ORGANISM: Chlamydomonas  
US-08-745-404-2

Query Match 34.5%; Score 44.5; DB 3; Length 516;  
Best Local Similarity 51.7%; Pred. No. 80;  
Matches 15; Conservative 4; Mismatches 7; Indels 3; Gaps 2;

Qy 1 SDALEALADQTDALQSEEAAYVK--ADNA 27  
:||||| : | :||||| | || |  
Db 369 ADAVEAAAEELTA-QAEAAANAKWEADKA 396

RESULT 3  
US-09-320-774-8  
; Sequence 8, Application US/09320774  
; Patent No. 6265545  
; GENERAL INFORMATION:  
; APPLICANT: Jarvik, Jonathan W.  
; TITLE OF INVENTION: READING FRAME INDEPENDENT EPITOPE  
; FILE OF INVENTION: TAGGING  
; NUMBER OF SEQUENCES: 47  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Harris Brotman  
; STREET: 202 Coast Blvd., Suite 111  
; CITY: La Jolla  
; STATE: California  
; COUNTRY: US  
; ZIP: 92037

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/320.774  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/762.106  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Brotman, Harris F.  
; REGISTRATION NUMBER: 35.461  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 654-2428  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 516 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; FRAGMENT TYPE: internal  
US-09-320-774-8

Query Match 34.5%; Score 44.5; DB 4; Length 516;  
Best Local Similarity 51.7%; Pred. No. 80;  
Matches 15; Conservative 4; Mismatches 7; Indels 3; Gaps 2;

Qy 1 SDALEALADQTDALQSEEAAYVK--ADNA 27  
:||||| : | :||||| | || |  
Db 369 ADAVEAAAEELTA-QAEAAANAKWEADKA 396

RESULT 4  
US-08-762-106-9  
; Sequence 9, Application US/08762106  
; Patent No. 5948677  
; GENERAL INFORMATION:  
; APPLICANT: Jarvik, Jonathan W.  
; TITLE OF INVENTION: READING FRAME INDEPENDENT EPITOPE  
; FILE OF INVENTION: TAGGING  
; NUMBER OF SEQUENCES: 47  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Harris Brotman  
; STREET: 202 Coast Blvd., Suite 111  
; CITY: La Jolla  
; STATE: California  
; COUNTRY: US  
; ZIP: 92037

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/762.106  
; FILING DATE: 09-DEC-1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Brotman, Harris F.  
; REGISTRATION NUMBER: 35.461  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 654-2428  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 527 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; FRAGMENT TYPE: internal  
US-08-762-106-9

Query Match 34.5%; Score 44.5; DB 2; Length 527;  
Best Local Similarity 51.7%; Pred. No. 82;  
Matches 15; Conservative 4; Mismatches 7; Indels 3; Gaps 2;

Qy 1 SDALEALADQTDALQSEEAAYVK--ADNA 27  
:||||| : | :||||| | || |  
Db 380 ADAVEAAAEELTA-QAEAAANAKWEADKA 407

RESULT 5  
US-09-320-774-9  
; Sequence 9, Application US/09320774  
; Patent No. 6265545  
; GENERAL INFORMATION:  
; APPLICANT: Jarvik, Jonathan W.  
; TITLE OF INVENTION: READING FRAME INDEPENDENT EPITOPE  
; FILE OF INVENTION: TAGGING  
; NUMBER OF SEQUENCES: 47  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Harris Brotman  
; STREET: 202 Coast Blvd., Suite 111  
; CITY: La Jolla  
; STATE: California  
; COUNTRY: US  
; ZIP: 92037

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible

GenCore version 5.1.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2002, 04:45:42 ; Search time 4.34235 Seconds  
(without alignments)  
157.503 Million cell updates/sec

Title: US-09-847-539a-6\_COPY\_59\_86

Perfect score: 129

Sequence: 1 SDALRALADQTDALQSEAAVVKADNAA 28

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 2442594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_AA.\*

- 1: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep.\*
- 2: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep.\*
- 3: /cgn2\_6/ptodata/2/1aa/6A\_COMB.pep.\*
- 4: /cgn2\_6/ptodata/2/1aa/6B\_COMB.pep.\*
- 5: /cgn2\_6/ptodata/2/1aa/PCTUS\_COMB.pep.\*
- 6: /cgn2\_6/ptodata/2/1aa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	44.5	34.5	516	2	US-08-762-106-8
2	44.5	34.5	516	2	US-08-745-404-2
3	44.5	34.5	516	4	US-09-320-774-8
4	44.5	34.5	527	2	US-08-762-106-9
5	44.5	34.5	527	4	US-09-320-774-9
6	44.5	34.5	552	3	US-08-745-404-3
7	44	34.1	344	4	US-08-961-083-192
8	44	34.1	514	2	US-08-960-022-14
9	43.5	33.7	382	1	US-08-049-2828-2
10	43.5	33.7	382	2	US-08-537-715-2
11	43.5	33.7	382	5	PCT-US94-04173-2
12	43.5	33.7	383	2	US-08-537-715-4
13	43.5	33.7	383	5	PCT-US94-04173-4
14	43.5	33.7	449	2	US-08-657-392-2
15	43.5	33.7	449	5	PCT-US94-02539-2
16	43.5	33.7	471	2	US-08-657-392-31
17	43.5	33.7	471	5	PCT-US94-02539-31
18	43.5	33.7	528	4	US-08-928-2138-8
19	43.5	33.7	711	1	US-08-235-838-7
20	43.5	33.7	711	2	US-08-465-4738-7
21	43	33.3	407	2	US-08-934-846-4
22	43	33.3	407	4	US-09-238-557-4
23	43	33.3	418	2	US-08-934-846-2
24	43	33.3	418	4	US-09-238-557-2
25	43	33.3	744	1	US-08-764-100-25
26	42.5	32.9	1611	2	US-08-804-227C-5
27	42	32.6	102	2	US-08-710-749-8

28 42 32.6 427 4 US-09-521-668B-18 Sequence 18, Appl  
29 42 32.6 427 4 US-09-521-668B-20 Sequence 20, Appl  
30 42 32.6 2890 4 US-09-413-814-67 Sequence 67, Appl  
31 42 32.6 3798 3 US-09-335-409-6 Sequence 6, Appl  
32 42 32.6 3798 4 US-09-568-102-6 Sequence 6, Appl  
33 42 32.6 3798 4 US-09-567-969-6 Sequence 6, Appl  
34 42 32.6 3798 4 US-09-568-480-6 Sequence 6, Appl  
35 42 32.6 3798 4 US-09-568-486-6 Sequence 6, Appl  
36 42 32.6 3798 4 US-09-568-472-6 Sequence 6, Appl  
37 41 31.8 173 3 US-08-669-408B-8 Sequence 8, Appl  
38 41 31.8 370 5 PCT-US93-01676A-4 Sequence 4, Appl  
39 41 31.8 395 1 US-07-876-280-9 Sequence 9, Appl  
40 41 31.8 395 1 US-08-049-783-6 Sequence 6, Appl  
41 41 31.8 395 1 US-08-316-301A-10 Sequence 10, Appl  
42 41 31.8 395 2 US-08-504-278-10 Sequence 10, Appl  
43 41 31.8 395 4 US-09-076-137-10 Sequence 10, Appl  
44 41 31.8 395 4 US-09-222-594-10 Sequence 10, Appl  
45 41 31.8 395 5 PCT-US92-03624-10 Sequence 10, Appl

## ALIGNMENTS

RESULT 1

US-08-762-106-8  
; Sequence 8, Application US/08762106  
; Patent No. 5948677  
; GENERAL INFORMATION:  
; APPLICANT: Jarvik, Jonathan W.  
; TITLE OF INVENTION: READING FRAME INDEPENDENT EPITOPE  
; NUMBER OF INVENTIONS: TAGGING  
; NUMBER OF SEQUENCES: 47  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Harris Brotman  
; STREET: 202 Coast Blvd., Suite 111  
; CITY: La Jolla  
; STATE: California  
; COUNTRY: US  
; ZIP: 92037  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION NUMBER: US/08/762,106  
; FILING DATE: 09-DEC-1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Brotman, Harris F.  
; REGISTRATION NUMBER: 35,461  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 654-2428  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 516 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; FRAGMENT TYPE: internal  
US-08-762-106-8

Query Match 34.5%; Score 44.5; DB 2: Length 516;  
Best Local Similarity 51.7%; Pred. No. 80;  
Matches 15; Conservative 4; Mismatches 7; Indels 3; Gaps 2;

QY 1 SDALRALADQTDALQSEAAVVK--ADNA 27

Db 369 ADAVEAAALTA-QAEAAARWEADKA 396

RESULT 2

US-08-745-404-2



```
/transl_table=11
/product="L-2,4-diaminobutyrate decarboxylase"
/protein_id="AG20873.1"
/db_xref="GI:10584266"
/translation="MEACDADVLVDVIGGAHGPYSGASYRLRLRLADTTALPAEGHPL
EDVLAARDVLNAVHPDETCVAHLOCPPMVPALAAEALLTASNOISDSFDOAPAA
TVEERLIADLTALVELTGAADGVVGGTGESNHQALLARDDYVETVPQSVREHGL
PAAQDLRLCSADAHFTAAQSAALLGIGEDAVYIPTGGRHMDATALREVVERLND
AKGHPAVVGTAGTDFGSDISLAVADAATHDLMVHVDAAFGGALAVSDQSERLS
GIDRADSVDFHKKFFQPIACGALLVADGESFELMNAAYLNPAGDAVPLNVAKST
RTTRFDALVYVAFRTLGRDIGALVDTLALADVAGLLRADPAFELACCEPTLNAY
TTRYREVRHAHICDQCPMADHVTEAREOLFAGTAVARTTVDDRAHVKETLMNPRT
TVADIOSVLVAFKGAATIEAADRTAPITTAAGDQ"
4376..6199
/gene="lucA"
/notes="VNG6212G"
4376..6199
/gene="lucA"
/notes="lucA"
/codon_start=1
/transl_table=11
/product="iron transport protein A"
/protein_id="AAG20874.1"
/db_xref="GI:10584267"
/translation="MTGVCADVADTPAAHAESATLHAFNLCLRETDAGEVVSRAVPD
GYPACQADRVRVVRPFPAAQSTTVFAPLRYASATGRHLFAVPAYARPGVPDAPAGTP
SPIDAGVASLARRELALOGDDPTAGRDGLERVLASQATATFRVAAORTPROHERLYD
TIVSFAQSQSLVYGHMHPTNSRTGIPPHQESRYAPELRAGFQLRVFAADPALVET
MAAGDASATEMLRSVADATVTPPESEATLVPVHPWQADYHAQPHVAAAIARDELRL
LGAFGPTFTYTSVRTLWSPDAPFMVKGSLVIEITNAERTNKRPELARGVAVSELLAT
GFGDAFADRFPAFIEDPAATLTLGDCAESGEAEVLRENPPAGDDTDTNVTVAALC
OQLDGPAAARVRIATREGRTESAVAREWFEQYLAHTVVRVWLYLFOQGVLEAH
QNTLVRLDDGRPTAGFYRDNQGYFFPASQAADVDMLPGVCERADITVCPDAIADER
LRYIVFNAGVGNALGAGGVVDERALLAVLRTELESIAEHEDPADSTLVLSALLTERR
VYKANLIRFRFEDELEADLETQSVYTAIENPLVTRVDQ"
6196..6906
/gene="lucB"
/notes="VNG6213G"
6196..6906
/gene="lucB"
/notes="lucB"
/codon_start=1
/transl_table=11
/product="iron transport protein B"
/protein_id="AAG20875.1"
/db_xref="GI:10584268"
/translation="MTAPLRITSDYAYRRPDQIDRTVAFRRVALDRDLGLRLHAWLT
SDHVTYVSLGAPLPAFHQAIGDLEDHVTAYVGCIDGVPMSYVRYVWVAAADPLADH
YDADNDQGIHILIGPPEYLGCGGTALLRAMTALAFRAHAATDRVVAEPDATNDAAIA
AFERAGFEPRETFREPHEKDATRVCVDRERFVGAFSPCAIAANGPPADGPVANHDDA
THRCPSGEGHNGSTLEADR"
6903..8336
/gene="hxyA"
/notes="VNG6214G"
6903..8336
/gene="hxyA"
/notes="hxyA"
/codon_start=1
/transl_table=11
/product="monooxygenase"
/protein_id="AAG20876.1"
/db_xref="GI:10584269"
/translation="MTFESVHDLVGVGLGPNFNLGLAALIEGSDVDVPLDAALFLOKPEFV
WHDGMILECTTLEVPFLADLVLTADTNPNYSYLNLYRHRDLRYEFYFYFTQIPREY
NDYCWQWASLEAPRFESRRVDITAIDDDATPANFRVTAVEPADARSATDDDDPHQ
YRARDLVCGSRPVVPPAFRDLPAFDVHTAAVREHRAACVADDSVAVYIGSQSAAE
VYHDLTRESTSDRSRVDWVTRSAGFPMEYSKLGLQHTPEYTDYVDLQPVKDE
LIPNDQDILYKGDPTSTRIYDTLYEQSIGGRDPNVMGTAIEVGAVEFTGGGRVVD
CHRRQDTGFATETDAVILGTGYHRPTPGFLSSLEPLRLARDDDODRLAVDREYGVGGDL
PGCLFQVQALETHGVGADPLGIGAYRNAVILDRDLADAPYVVDRTVTPQDFAPEAYA
SHAPGGRVDRDDPTTPPAAGQPNDDHQ"
8362..10188
/gene="lucC"
```

```
/note="VNG6216G"
8362..10188
/gene="lucC"
/notes="lucC"
/codon_start=1
/transl_table=11
/product="iron transport protein C"
/protein_id="AAG20877.1"
/db_xref="GI:10584270"
/translation="MNTYEDTLQDAVTADTWATVERSLIAKLRLRETYBELIAPERTT
DTRDGRYRLVLPAAADADTVRYRFEAEERLMDTLRVFPOTIERADADADSWATATDP
IRFLDLADTLQTLTGLHIREYRNTLLADAHIEARSAPESFADCSVAEYSESEMTG
HKPIFNKRGRLGNDVEDYDYAPAPERQTTIGNLAVRTDRVTEETVSGLTPTGLRGE
LGDQDDFMATLADRGIDPQETTLPLPVHNYQWETTITLSIASELADAIVPLNGADS
VLPQGSVRTLFNVDAPEKHNVKLPKIRNTLVWRGLPGDRTAAAPRVTEITDIRDGD
PFLRDDCELVLFEVAGVADYRHPAFSELDNAYQYDELLGAVMRESITDAIPDSQOAV
TLAALLHEADGTAVISERFVSPLSELSAWLTELFDLPLLLHLYLVRYGTVPFSPHGEN
TIVLDEEVPTRLGVKDFVDVNI SDQPLPELADLADLRRLVLRSEPTNGLCQFYFC
GLFVGVRVYLSVDLTAHHCYDETEFTQVRDAITAVQAEFPELSDREFTDLLQPTFT
KICLNNRNLTEEGYSDTASRPAAAEHGTVTNALHEVAGDAE"
complement(10189..10842)
/gene="trkA3"
/notes="VNG6218G"
complement(10189..10842)
/gene="trkA3"
/notes="trkA3"
/codon_start=1
/transl_table=11
/product="Trk potassium uptake system protein"
/protein_id="AAG20878.1"
/db_xref="GI:10584271"
Alignment Scores:
Pred. No.: 674 Length: 12423
Score: 51.00 Matches: 12
Percent Similarity: 65.38% Conservative: 5
Best Local Similarity: 46.15% Mismatches: 9
Query Match: 39.53% Indels: 0
DB: 1 Gaps: 0
US-09-847-539A-6_COPY_59_86 (1-28) x AE005155 (1-12423)
QY 2 AspAlaLeuGluaLaLeuAlaAspGlnThrAspAlaLeuGlnSerGluiGluAlaVal 21
Db 3981 GACGGACTCGGGCGGCGTGGTGAACGCTCGCGCTCCGCCGACGACGCTGGGGGCGCTG 4040
QY 22 VailysAlaAspAsnAla 27
Db 4041 CTGCGTGCTCACCCGGCG 4058
RESULT 40
AF072878 14941 bp DNA linear INV 21-DEC-1999
LOCUS Tetrahymena thermophila ciliary outer arm dynein beta heavy chain
DEFINITION (DYH4) gene, complete cds.
ACCESSION AF072878
VERSION AF072878.2 GI:5209335
KEYWORDS SOURCE Tetrahymena thermophila.
ORGANISM Tetrahymena thermophila
Eukaryota; Alveolata; Ciliophora; Oligohymenophorea;
Hymenostomatida; Tetrahymenina; Tetrahymena.
REFERENCE 1 (bases 1 to 14941)
AUTHORS Lee,S., Wisniewski,J.C., Dentler,W.L. and Asai,D.J.
TITLE Gene knockouts reveal separate functions for two cytoplasmic
dyneins in Tetrahymena thermophila
JOURNAL Mol. Biol. Cell 10 (3), 771-784 (1999)
MEDLINE 99169020
PUBMED 10069817
REFERENCE 2 (bases 1 to 14941)
AUTHORS Lincoln,L.M., Gibson,T.M., Asai,D.J. and Forney,J.D.
TITLE A gene knockout reveals that dynein beta heavy chain is required in
Tetrahymena thermophila
```





```
transcript"
/codon_start=1
/db_xref="FLYBASE:FBan0001030"
/db_xref="FLYBASE:FBgn0000339"
/protein_id="AAF54082.1"
/db_xref="GI:7298875"
/translation="MDPCFAMSSYQFVNSLASCYPOQNPQNHGPGAGSSAGSGG
GAGSGGVVPSGCTNGGGSAGATPGANDYFPAAAYTPNLVPTPOAHYANQAAYG
CGQNDVDYITQLQRLULLQQQQQQQHAHAHAAYAAQQQQLAQQQHPQQQQQ
QOANISCIYANDVPTPGSGGGVSNNNNSNNNSQSLASQDLSTRDISPK
LSPSSVBSKSLKGVIGSLAAAGLNNHNSGSGGPGNVNVPWHSPPGG
DSDESDEAGNSGNSGKKNPPIYPMKRVHLGCTVNVANGETKRTORTSVTRY
QTLKEKEFENRVLTRRRRIEIAHALCLTERQIKWFQNRMRMKKEKKMASNIVP
YHMPGYCHPIHQFDIHPQFAHLS"
join(86526..86589,86623..87370,87518..88310)
/gene="ftz"
/note="Nucleotide sequence of the Celera sequence differs
from the published sequence for this transcript."
/product="CT6611"
/db_xref="FLYBASE:FBan0002047"
/db_xref="FLYBASE:FBgn0001077"
<86526..>88310
/gene="ftz"
/map="84B1-84B1"
/db_xref="FLYBASE:FBan0002047"
/db_xref="FLYBASE:FBgn0001077"
join(86623..87370,87518..88002)
/gene="ftz"
/note="ftz gene product; Nucleotide sequence of the Celera
sequence differs from the published sequence for this
transcript."
/codon_start=1
/db_xref="FLYBASE:FBan0002047"
/db_xref="FLYBASE:FBgn0001077"
/protein_id="AAF54081.1"
/db_xref="GI:7298874"
/translation="MATNNSQSHYSYADNMNMWYHPSLPTYYDNSGNAYQNT
SNYQYTPQESYSSCYTYNNGQYTTQTPVPQTPPPKATRKKAEDDAASIIAAV
BERSTLRALTNPVKIKLYTPDYFTVEQVKAPAVSTKVTASPSPDYQYVTV
TPSASEVDYLDVTSPOQKNGDFATPPTTSLPPLLEGISTPSPGKSSS
AVSOEIHRIYTPAGAGDNFKNSHIEETLASDCKDSKRTQTYRYOTLEKEKEFH
RYITRRRRIDIANLSEROIKIWFQNRMRMKSKDRTLDSPHCAGAGYAMLPPLLE
ATSVATTGAPSVPMYHHHQTAAYPAYSHSHSGYLLNDYFQQQTHQYDAYPQ
YHQCSYQHQDLYLHS"
complement(join(118482..120767,130742..130967,
131102..131140,131519..132295,133615..1393704,
221238..221367,221470..222034,222070..222260,
222819..223415,223431..223587))
/gene="Antp"
/product="CT30093"
/db_xref="FLYBASE:FBan0001028"
/db_xref="FLYBASE:FBgn0000095"
complement(<118482..>223587)
/gene="Antp"
/note="CG1028"
/map="84B1-84B4"
/db_xref="FLYBASE:FBan0001028"
/db_xref="FLYBASE:FBgn0000095"
complement(join(120517..120767,130742..130967,
131102..131140,131519..132139))
/gene="Antp"
/note="Antp gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0001028"
/db_xref="FLYBASE:FBgn0000095"
/protein_id="AAG22205.1"
/db_xref="GI:10727115"
/translation="WTMTNNCESMTSYFTNSYMGADMHHGHYPGNGVTDLDAQONHH
YISQANHGNNPFRFPPTDRMPYNGQMQQQQQHQQYVSRPSPSSQVGGVMPQAT
NGLGVPGQQQQQQQQSONQQQAQQQLAQQLQVTTQVTHPQQQQQQQVYVA
SCKLQAAYVGLGMYEGGSPPLVDQMGGHNNNAQMTLPHHMHPQALQGYTDVGPV
TEVQHNNHNMGMYYQQSGVPPVGAPPQGMHQGGPPQMHQGHPPQHTPPSQNPNSQS
```

```
SCMPSLPYPMRSGKQCKRGRQTYTRYOTLEKEFENRVLTRRRRIEIAHAL
CLTERQIKWFQNRMRMKKEKKTKGEPGSGGEGDEITPPNSPQ"
join(275368..275682,275741..275774,276161..276630,
276896..277114,277173..277294,277348..>277531)
/gene="Sodh-1"
/note="Nucleotide sequence of the Celera sequence differs
from the published sequence for this transcript."
/product="CT6209"
/db_xref="FLYBASE:FBan0001982"
/db_xref="FLYBASE:FBgn0024289"

Alignment Scores:
Pred. No.:      Length:      309357
Score:          51.50        Matches:      13
Percent Similarity: 65.38%    Conservative:  4
Best Local Similarity: 50.00%  Mismatches:    8
Query Match:      39.92%     Indels:         1
DB:               3          Gaps:         1

US-09-847-539a-6_COPY_59_86 (1-28) x AE003673 (1-309357)
QY  2  AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuSerGluGluAlaIaVal 21
|||||:::|||||:::|||||:::|||||:::
Db 58507 GAGCGTTTAAAGCGTTTCTTCACAGACTGAT---TTGAGCCATCCGATGATATT 58563
QY  22 ValLysAlaAspAla 27
|||||:::|||||:::|||||:::|||||:::
Db 58564 TTTAAGGCAGACAATGCC 58581

RESULT 38
AY061146/c
LOCUS AY061146
DEFINITION Drosophila melanogaster LD12115 full length cDNA.
ACCESSION AY061146
VERSION AY061146.1 GI:16768949
KEYWORDS FLI CDNA.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Mecoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 2151)
AUTHORS Stapleton,M., Brokstein,P., Hong,L., Agbayani,A., Carlson,J.,
Champe,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E., George,R.,
Gonzalez,M., Guarin,H., Li,P., Liao,G., Miranda,A., Mungall,C.J.,
Nunoo,J., Pacleb,J., Paragas,V., Park,S., Phouanavong,S., Wan,K.,
Yu,C., Lewis,S.E., Rubin,G.M. and Celniker,S.
Direct Submission
Submitted (30-OCT-2001) Berkeley Drosophila Genome Project,
Lawrence Berkeley National Laboratory, One Cyclotron Road,
Berkeley, CA 94720, USA
Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory
Berkeley, CA 94720
This clone was sequenced as part of a high-throughput process to
sequence clones from Drosophila Gene Collection 1 (Rubin et al.,
Science 2000). The sequence has been subjected to integrity checks
for sequence accuracy, presence of a polyA tail and contiguity
within 100 kb in the genome. Thus we believe the sequence to
reflect accurately this particular cDNA clone. However, there are
artifacts associated with the generation of cDNA clones that may
have not been detected in our initial analyses such as internal
priming, priming from contaminating genomic DNA, retained introns
due to reverse transcription of unspliced precursor RNAs, and
reverse transcriptase errors that result in single base changes.
For further information about this sequence, including its location
and relationship to other sequences, please visit our Web site
(http://fruitfly.berkeley.edu) or send email to
cdna@fruitfly.berkeley.edu.
Location/Qualifiers
1..2151
/organism="Drosophila melanogaster"
```

```

/chromosome="3R"
/map="84A-84B"
/clone="BACR32J03 (D1350)"
/clone_lib="RPCI-98 (Roswell Park Cancer Institute
Drosophila melanogaster BAC library, partial EcoRI in
pBACE3.6)"
BASE COUNT 55773 a 39290 c 38978 g 56601 t
ORIGIN

Alignment Scores:
Pred. No.: 1e+04 Length: 190642
Score: 51.50 Matches: 13
Percent Similarity: 65.38% Conservative: 4
Best Local Similarity: 50.00% Mismatches: 8
Query Match: 39.92% Indels: 1
DB: 3 Gaps: 1

US-09-847-539A-6_COPY_59_86 (1-28) x AC095015 (1-190642)

Qy 2 AspalateuGluAlaLeuAlaAspGlnThrAspalateuGlnSerGluGluAlaAlaVal 21
|||||:::|||| ::| ||||||| ||| ::| ::|
Db 3240 GACGCTTTAAAGCGTTTCTTCACAGACTGAT---TTGAGCCATCCGATGAGTGATATT 3296

Qy 22 vallysalaaspanAla 27
|||||:::|||||
Db 3297 TTTAAGGCAGACAATGCC 3314

RESULT 37
AE003673
LOCUS
DEFINITION
Drosophila melanogaster genomic scaffold 14200001386040 section 3
of 5, complete sequence.
ACCESSION
AE003673 AE002699
VERSION
AE003673.2 GI:10727114
KEYWORDS
HTG.
SOURCE
fruit fly.
ORGANISM
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 30357)
Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
Ananides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galie,R.F.,
George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
Sutton,G.G., Rogers,J.R., Vandeil,M.D., Zhang,Q., Chen,L.X.,
Brandon,R.C., Watters,J.H., Blazej,R.G., Champe,M., Pfeiffer,B.D.,
Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor
Miklos,G.L., Abril,J.F., Agbayani,A., An,H.J.,
Andrews-Pfannkoch,C., Baldwin,D., Ballew,R.M., Basu,A.,
Baxendale,J., Beckstead,O., Bergelson,D., Bessie,E.M., Beeson,K.Y.,
Benos,P.V., Bertram,B.P., Bhandari,D., Bolshakov,S., Borkova,D.,
Botchan,M.R., Bouck,J., Brokstein,P., Brottier,P., Burtis,K.C.,
Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferrieta,S.,
Fleischmann,W., Fosler,C., Gabrielian,A.E., Garg,N.S.,
Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,
Guan,P., Harris,M., Harris,N.L., Harvey,D., Helman,T.J.,
Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,
Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,
Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattle,B., McIntosh,T.C.,
McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Moberly,C.,
Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
Nusskern,D.R., Pacleb,J.M., Palazzolo,M., Pittman,G.S., Pan,S.,
Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,
Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I.,
Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,
Stapleton,M., Strong,R., Sun,E., Swirskas,R., Tector,C., Turner,R.,

```

```

ACCESSION AC095014
VERSION AC095014.1 GI:15624856
KEYWORDS HTG.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 170801)
REFERENCE 1
AUTHORS Celniker,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H.,
Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C.,
Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,
Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M.,
Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,
Ferreira,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A.,
Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattel,B., Moshrefi,A.,
McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,
Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,
Phouanavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
Stapleton,M., Strong,R., Svirskas,R., Tector,C., Williams,S.M.,
Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
Unpublished
2 (bases 1 to 170801)
REFERENCE 2
AUTHORS Celniker,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H.,
Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C.,
Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,
Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M.,
Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,
Ferreira,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A.,
Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattel,B., Moshrefi,A.,
McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,
Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,
Phouanavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
Stapleton,M., Strong,R., Svirskas,R., Tector,C., Williams,S.M.,
Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
Direct Submission
Submitted (15-SEP-2001) Berkeley Drosophila Genome Project, MS
64-121, Lawrence Berkeley National Laboratory, One Cyclotron Road,
Berkeley, CA 94720, US
Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720
This sequence was assembled using end sequences from a whole genome
shotgun and from subclones of this BAC and its neighboring clones.
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive Web site (http://www.fruitfly.org/sequence/) or send email
to bdg@fruitfly.berkeley.edu.
Location/Qualifiers
1. 170801
/organism="Drosophila melanogaster"
/strain="y: cn bw sp"
/db_xref="taxon:7227"
/chromosome="3R"
/map="83F-84A"
/clone.lib="BAC19D23 (D1349)"
Drosophila melanogaster BAC library, partial EcoRI in
PRACE3.6"
BASE COUNT 49244 a 35184 c 35797 g 50576 t
ORIGIN
Alignment Scores:
Pred. No.: 8,91e+03 Length: 170801
Score: 51.50 Matches: 13
Percent Similarity: 65.38% Conservative: 4
Best Local Similarity: 50.00% Mismatches: 8
Query Match: 39.92% Indels: 1
DB: 3 Caps: 1
US-09-847-539A-6_COPY_59_86 (1-28) x AC095014 (1-170801)
Qy 2 AspaLaLeuGluaLaLeuAlaAspGlnThrAspAlaLeuGlnSerGlUGluAlaVal 11
|||||:||||| :|: |||||:| ||| :|: :|:
Db 170223 GACGCTTTAAAGCGCTTTCTTCACAGACTGAT---TTGAGCCATCCGATCAGTGATATT 170279
Qy 22 VallysAlaAspAsnAla 27
|||||:|||||:|
Db 170280 TTTAAGGCAGACAAATGCC 170297
RESULT 36
AC095015
LOCUS Drosophila melanogaster, chromosome 3R, region 84A-84B, BAC clone
DEFINITION BACR3J03, complete sequence.
ACCESSION AC095015
VERSION AC095015.1 GI:15624857
KEYWORDS HTG.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 190642)
REFERENCE 1
AUTHORS Celniker,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H.,
Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C.,
Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,
Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M.,
Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,
Ferreira,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A.,
Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattel,B., Moshrefi,A.,
McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,
Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,
Phouanavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
Stapleton,M., Strong,R., Svirskas,R., Tector,C., Williams,S.M.,
Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
Unpublished
2 (bases 1 to 190642)
REFERENCE 2
AUTHORS Celniker,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H.,
Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C.,
Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,
Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M.,
Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,
Ferreira,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A.,
Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattel,B., Moshrefi,A.,
McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,
Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,
Phouanavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
Stapleton,M., Strong,R., Svirskas,R., Tector,C., Williams,S.M.,
Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
Direct Submission
Submitted (15-SEP-2001) Berkeley Drosophila Genome Project, MS
64-121, Lawrence Berkeley National Laboratory, One Cyclotron Road,
Berkeley, CA 94720, US
Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720
This sequence was assembled using end sequences from a whole genome
shotgun and from subclones of this BAC and its neighboring clones.
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive Web site (http://www.fruitfly.org/sequence/) or send email
to bdg@fruitfly.berkeley.edu.
Location/Qualifiers
1. 190642
/organism="Drosophila melanogaster"
/strain="y: cn bw sp"
/db_xref="taxon:7227"

```

```
/db_xref="GI:14575550"
/translation="MSDVTDAVARLVDRHGTPVPALVLASLAALATAGFHTRWRRHG
CAPPADTGARPPAPAGTGPAGSEGGSGAGGESALWRMTTVRDAPGSL
AVLCAALADRIDILSLQTHPLTCTGVDFLLRAPGELSAELAENVYTAGGARTWTE
RADAHLDVADTRVLGLAARVALDAELPLALRQLLGRCTIRSLPAPADDDGAGARAAA
AVPPEGALEDTVLRLRAPEGGVSVVERPHLPFTFAEFARALVELDARLPRVPRGR
DVLTLPEGSDDTVRRADTRDYPAAKAMHERCSARTLGMRYHGPVGDADRYLHLLSPR
FGTLAVTAGSLVGLHLLMDGDETEVALLVEDGMQRGIGGELLARLVAMAADAG
CESVVAITQASNTGMTVMAMRGLSLPLDYQIEEGLTVTARLDPAQAARALPQAEIRG
RE"
```

## misc\_feature

```
5557..5766
/gene="SCBAC36F5.04"
/note="p1fam match to entry PF01842 ACT, ACT domain, score
17.60, E-value 0.29"
6310..6516
/gene="SCBAC36F5.04"
/note="p1fam match to entry PF00583 Acetyltransf,
Acetyltransferase (GNAT) family, score 36.20, E-value
7.7e-07"
```

## gene

```
complement(6676..7899) .
```

## CDS

```
/gene="SCBAC36F5.05c"
complement(6676..7899)
/gene="SCBAC36F5.05c"
/note="SCBAC36F5.05c, possible lyase, len: 407 aa; similar
to SW:MTB_HAEIN (EMBL:U32694) Haemophilus influenzae
cystathionine gamma-synthase (EC 4.2.99.9) MetB or HI0086,
369 aa; fasta scores: opt: 625 Z-score: 685.3 bits: 135.6
E(): 1.5e-30; 34.127% identity in 378 aa overlap. Contains
p1fam match to entry PF01053 Cys_Met_PP, Cys/Met
metabolism PLP-dependent enzyme and match to Prosite entry
PS01050 Uncharacterized protein family UPF0031 signature
2"
```

```
/codon_start=1
/transl_table=11
/product="putative lyase"
/protein_id="CAC4284i.1"
/db_xref="GI:14575551"
/translation="MDSARTHTSPDDVRFRTAPAGALATEAVHAGRDDLAQGLHAP
PIDLSTYPSYDSRGEAARIDAFAATGAEPDGPVYGRGNPTVAFETALARLEGTE
ADAFASGMAALSALLVLRGSMGLRHVVAVRPLYGCSDBLLTAGLLGSEVTWTPAGV
AVALRPDTGLVLVESPNPTLAEDVLRAVAHACGVPVLLVDNTATPVLORPAEHGAR
LVLSATYLLGGHGDVLGVVACDEEFAGRLRQVRFPATGGVHLPLAGYLLRGLSTLP
```

## Alignment Scores:

```
Pred. No.: 1.55e+03 Length: 32608
Score: 51.50 Matches: 17
Percent Similarity: 50.00% Conservative: 3
Best Local Similarity: 42.50% Mismatches: 7
Query Match: 39.92% Indels: 13
DB: 1 Gaps: 2
```

US-09-847-539A-6\_COPY\_59\_86 (1-28) x SCBAC36F5 (1-32608)

```
QY 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThr----- 11
Db 13175 GCGCGCGCGCGCGCGCGCGTCTCGTGCAGACACCGCCGAGTGTCAACGACGGCTC 13116
QY. 12 AspAlaLeuGlnSerGluGluAlaAlaVal-----ValLysAlaAspAsnAla 27
Db 13115 GACGCTGCGCAGTCGGAGACCGCGCGGTCCGTGCGGTCTCGCGGAGCCGACCGCC 13056
```

## RESULT 33

```
AC012724
LOCUS Drosophila melanogaster, 35031 bp DNA linear HTG 03-NOV-1999
DEFINITION pieces.
```

## ACCESSION

AC012724

## VERSION

AC012724.1 GI:6223583

## KEYWORDS

HTG; HTGS\_PHASE2.

## SOURCE

fruit fly.

## ORGANISM

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 35031)

AUTHORS Adams M. and Venter J.C.

TITLE Direct Submission

JOURNAL Submitted (03-NOV-1999) Celera Genomics, 45 West Gude Drive,

Rockville, MD, USA

COMMENT This sequence was identified as CDM:10209782 by the submitter.

For further information on this sequence you may e-mail to

fly@celera.com.

\* NOTE: This is a 'working draft' sequence.

\* This sequence will be replaced

\* by the finished sequence as soon as it is available and

\* the accession number will be preserved.

FEATURES

1..35031

Location/Qualifiers

/organism="Drosophila melanogaster"

/db\_xref="taxon:7227"

BASE COUNT 9943 a 7273 c 7242 g 10573 t

ORIGIN

Alignment Scores:

Pred. No.: 1.67e+03 Length: 35031

Score: 51.50 Matches: 13

Percent Similarity: 65.38% Conservative: 4

Best Local Similarity: 50.00% Mismatches: 8

Query Match: 39.92% Indels: 1

DB: 2 Gaps: 1

US-09-847-539A-6\_COPY\_59\_86 (1-28) x AC012724 (1-35031)

QY 2 AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluAlaAlaVal 21

Db 13803 GACGCTTTAAAGCGTTTCTTCACAGACTGAT---TTGAGCCATCCGATGATGATAT 13859

QY 22 ValLysAlaAspAsnAla 27

Db 13860 TTTAAGGCGAGACAATGCC 13877

RESULT 34

AE001572.2/c

WPCOMMENT

Sequence split into 5 fragments LOCUS AE001572 Accession AE001572

Fragment Name Begin End

AE001572\_0 1 110000

AE001572\_1 100001 210000

AE001572\_2 200001 310000

AE001572\_3 300001 410000

AE001572\_4 400001 429825

Continuation (3 of 5) of AE001572 from base 200001 (AE001572 Drosophila melanogaster

Alignment Scores:

Pred. No.: 5.6e+03 Length: 110000

Score: 51.50 Matches: 13

Percent Similarity: 65.38% Conservative: 4

Best Local Similarity: 50.00% Mismatches: 8

Query Match: 39.92% Indels: 1

DB: 3 Gaps: 1

US-09-847-539A-6\_COPY\_59\_86 (1-28) x AE001572.2 (1-110000)

QY 2 AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluAlaAlaVal 21

Db 20551 GACGCTTTAAAGCGTTTCTTCACAGACTGAT---TTGAGCCATCCGATGATGATAT 20495

QY 22 ValLysAlaAspAsnAla 27

Db 20494 TTTAAGGCGAGACAATGCC 20477

RESULT 35

AC095014

LOCUS AC095014

DEFINITION Drosophila melanogaster, 170801 bp DNA linear INV 15-SEP-2001

BACR19D23, complete sequence.

Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomycetes.

1 (bases 1 to 32608)  
 Redenbach, M., Kieser, H.M., Denapalite, D., Eichner, A., Cullum, J., Kinashi, H. and Hopwood, D.A.  
 A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb Streptomycetes coelicolor A3(2) chromosome  
 Mol. Microbiol. 21 (1), 77-96 (1996)  
 97000351

2 (bases 1 to 32608)  
 Saunders, D. and Harris, D.  
 Submitted (27-JUN-2001) Streptomycetes coelicolor sequencing project, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge  
 CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof. David A. Hopwood, [3] John Innes Centre, Norwich Research Park, Colney, Norwich, Norfolk NR4 7UH, UK

Notes:  
 Streptomycetes coelicolor sequencing at The Sanger Centre is funded by the BBSRC and Beowulf Genomics  
 Details of S. coelicolor sequencing at the Sanger Centre are available on the World Wide Web.  
 {URL: <http://www.sanger.ac.uk/Projects/s.coelicolor/>} CDS are numbered using the following system eg SC7B7.0lc. SC (S. coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary strand).  
 The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous. The length in codons is given for each CDS.  
 Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database. The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon. Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4768-4778(1994)) and the FramePlot program of Bibb et al., Gene 30:157-66(1984) as implemented at <http://www.nih.go.jp/jun/cgi-bin/frameplot.pl>. CAUTION: We may not have predicted the correct initiation/translation. Where possible we choose an initiation codon (atg, gtg, ttg or att) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.  
 IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions. Cosmid BAC36F5 overlaps cosmid 2StG18 and cosmid 2StG61.  
 Location/Qualifiers  
 1..32608  
 /organism="Streptomycetes coelicolor"  
 /db\_xref="taxon:1902"  
 1..99  
 /gene="SCBAC36F5.01"  
 /note="nominal overlap with Streptomycetes coelicolor cosmid 2StG18"

misc\_feature  
 1..99  
 /gene="SCBAC36F5.01"  
 /note="nominal overlap with Streptomycetes coelicolor cosmid 2StG18"

source  
 1..32608  
 /organism="Streptomycetes coelicolor A3(2)"  
 /strain="A3(2)"  
 /db\_xref="taxon:100226"  
 /clone="cosmid BAC36F5"  
 1..891  
 /gene="SCBAC36F5.01"  
 <1..891  
 /gene="SCBAC36F5.01"  
 /note="SCBAC36F5.01, possible secreted phosphatase (fragment), len: >296 aa; similar to C-terminal region of SW:PPBD\_BACSU (EMBL:U49060) Bacillus subtilis alkaline phosphatase D precursor (EC 3.1.3.1) PhOD, 556 aa; fasta scores: opt: 277 Z-score: 319.4 bits: 68.0 E(): 3.5e-10;

30.258% identity in 271 aa overlap"  
 /codon\_start=1  
 /transl\_table=11  
 /product="putative secreted phosphatase (fragment)"  
 /protein\_id="CAC42837.1"  
 /db\_xref="GI:14575547"  
 /translation="LDENLRFAQVPSIVQWDDHEVRNNYVPGQVLDYDDRYTEKRVDYLAARARAFAPYFVSSLRFCAREERVRVHOGPLLDVFLDMRTVYRNANSPCEQAVDPQGLRGLEQLEWLKRELARSRAVKVIADMPGLVYVDPDTTEGAPHEIALAQSDPGAPLRGELQIAELLRRVKHRRITGTWLTADVHTTSQAQHYQPSRAAFAGDFPEPWFVSGPLSAGAFPASBELDGTGTFGPFRVFKAPTASNSPAGGYQFFGEVDIDGGTAEMTVRLREQDGTVLFTKVLRRPGVQVQ"  
 1042..2445  
 /gene="SCBAC36F5.02"  
 1042..2445  
 /gene="SCBAC36F5.02"  
 /note="SCBAC36F5.02, unknown, len: 467 aa"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein SCBAC36F5.02"  
 /protein\_id="CAC42838.1"  
 /db\_xref="GI:14575548"  
 /translation="MTRSLQAVAYRRPSVLESAGGQHLGLETSRGATPAGAVDHPRFAGFLTAPOKASTALLAVADVAARYQQQLRASLDPPVVTGSGDLRFESFGCGGYVARDLVEAGLDGGEVGHGTTNDVNNPLREALSRIGADDPHLRLVGPFEIAVTTLDGPVVEKKVPLPDRMLRGPAFAOVIAAGPDLRAELDAEAAREVRLALPGCRSAATLRLVVPAGRSRLPTTRAVPGAVCLPGPERLVALQRLVLAHATALRYLTPGAVASGAATASAEVLPAGMRLTITLSPDAARFGSGEGGVLDALADEAADAELISVLLAWEPRIDTADLAASGLTAERVRAALRLGTSGRVYDGTAEAYFHRELPYDAERVERINRLRSARALVAAAGVVDGPIGTVTATDGHVHRVREAGVLSCLMWAKYRGGRGPCXKHALAVRMVRRGSGESGAVRIDGGAR"  
 2454..5177  
 /gene="SCBAC36F5.03"  
 2454..5177  
 /gene="SCBAC36F5.03"  
 /note="SCBAC36F5.03, unknown, len: 907 aa"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein SCBAC36F5.03"  
 /protein\_id="CAC42839.1"  
 /db\_xref="GI:14575549"  
 /translation="MASAVATPAATATATATATGTLVAARAGTAEVALLDGMTOAORRAAPELKAVRQELRTARNSADSRRAHPALHAAGACOTGAANVLAANVLAAMDWRMSQASPAVLLHVLGDRDSCWLDVAHRAQRPAAASNVSELMAGLVRLSGCPVPTEAVVQGNLTHVGGQWGGTVCDDLRRDPLHAEVLAALFRTEGVTQLGSLSGDGPSEWSTA LARLTGEGVLDRRSVVDACVARLLRGAPAEIRVFLRLDKDLALTRREERRRAADWTA LADAMPTVAASHAQSVLGAELSGPRLAQVTEAVLFEKTKLVRAQLVLLGKVLTKRNSAAGELLPAARAFGHEDSDAQERALKLVVERHYGVKVAAPAAVEAEIEISALLASGSDALRAARVTLGATPPDSVPVAHEEVLPPQPPVRLAPAPAAVEAEIEISALLASGSDVAVFERALDGLVRHAILDRDALLLEALPPVVARWMAEGEPHGRSDDFARRHELPD DARAFDLLLLATLYKVTDTLHRVYVHDLGTLNSCAINSLARPEARVWEVNAIRLTPD QPFLLATPTWTGLLEPGELVDLDYRAGARVSAVDFGQALLRVDNRNRPALAAAEAAALGTEGERLRLRLTERAAPSVTRTADTIRLVEGVEYDLPRAEFEPFALRLLRGPNVPSDHWYCSHWSGMARPHWFALLPERPELVLAARLLRDVSEAAVHDDQGTAAAVLPHLADADGAGAVHL5VAYGLGARHAEDRLAAYDALLVLAARGVLEACRIGALDLCQLVRRGAVKPARLADAVRTAAATGANATVWAYLRQVLPVLLADLSTGCATVSSARGLE LLYAAVACBARTGEQSGSLPHLSGVADRRGTSLVLTQARRLREALAATAAAAA"  
 5293..6651  
 /gene="SCBAC36F5.04"  
 5293..6651  
 /gene="SCBAC36F5.04"  
 /note="SCBAC36F5.04, hypothetical protein, len: 452 aa; C-terminal part similar to TR:09FC64 (EMBL:AL391515) Streptomycetes coelicolor putative acetyltransferase SC4B10.23, 163 aa; fasta scores: opt: 287 Z-score: 300.0 bits: 63.3 E(): 4.2e-09; 42.857% identity in 133 aa overlap. Contains Pfam match to entry PF01842 ACT, ACT domain and PF00593 Acetyltransf, Acetyltransferase (GNAT) family"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="CAC42840.1"

1









Db 117572	CTGGAAGCTCTGTTCAAGCAGGTGGAGTCTCTCCAAAGCGCAGTGTCTCTTGTGGCA	117631
QY	24	Alaaspaen 26
Db 117632	GCCGACAAT	117640
RESULT 27		
AE003619		
LOCUS		
DEFINITION	Drosophila melanogaster genomic scaffold 142000013386055 section 12 of 63, complete sequence.	
ACCESSION	AE003619	AE002690
VERSION	AE003619.2	GI:10728652
KEYWORDS	HTG.	
SOURCE	fruit fly.	
ORGANISM	Drosophila melanogaster	
REFERENCE	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.	
AUTHORS	Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D., Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F., George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N., Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X., Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D., Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor Miklos G.L., Abri J.F., Agbayani A., An H.J., Andrews-Pfannkuch C., Baldwin D., Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M., Beeson K.Y., Benos P.V., Berman B.P., Bhattacharya D., Bolshakov S., Borokova D., Botchan M.R., Bouck J., Brokstein P., Brottier P., Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I., Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P., de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M., Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P., Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W., Folsler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K., Glodek A., Gong F., Gorrell J.J., Guan P., Harris M., Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Hock J., Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C., Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A., Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z., Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X., Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D., Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A., Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L., Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M., Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G., Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H., Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T., Spier E., Spradling A.C., Stapleton M., Strong R., Sun E., Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X., Wang Z.Y., Wassarman D.A., Weinstock G.M., Weissbach J., Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao O.A., Ye J., Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L., Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O., Gibbs R.A., Myers E.W., Rubin G.M. and Venter J.C.	
TITLE	The genome sequence of Drosophila melanogaster	
JOURNAL	Science	287 (5461), 2185-2195 (2000)
MEDLINE	20196006	
REFERENCE	2 (bases 1 to 281993)	
AUTHORS	Adams, M.D., Celniker, S.E., Gibbs, R.A., Rubin, G.M. and Venter, C.J.	
TITLE	Direct Submission	
JOURNAL	Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive, Rockville, MD, USA	
COMMENT	On Oct 9, 2000 this sequence version replaced gi:7297302.	
FEATURES	Location/Qualifiers	
SOURCE	1. 281993	
	/organism="Drosophila melanogaster"	
	/db_xref="taxon:7227"	
	/chromosome="2L"	
	join(421..927,1513..1971,2323..2899,2959..3222,3299..3351,	

\* the accession number will be preserved.

```

FEATURES
  source
    1. .36993
    /organism="Drosophila melanogaster"
    /db_xref="taxon:7227"
  BASE COUNT    9781 a  8675 c  8725 g  9812 t
  ORIGIN
    1.48e+03      Length:      36993
    Score:        52.00      Matches:      11
    Percent Similarity: 65.22%      Conservative: 4
    Best Local Similarity: 47.83%      Mismatches: 8
    Query Match:    40.31%      Indels:      0
    DB:             2          Gaps:        0

  US-09-847-539a-6_COPY_59_86 (1-28) x AC018007 (1-36993)

  QY    4  LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVallys 23
  Db    30206 CTGGAAGCTCTGTTCAAGCAGGTGGAGTCTCTCCAGCGCCGAGTTCCTCTCTGTGGCA 30265

  QY    24  AlaAspAsn 26
  Db    30266 GCCGACAAT 30274

  RESULT 25
  AC011910
  LOCUS
  DEFINITION
    Drosophila melanogaster clone RPC198-2017, *** SEQUENCING IN
  PROGRESS ***, 39 unordered pieces.
  AC011910
  VERSION
  AC011910.1 GI:6056126
  KEYWORDS
  HTG; HTGS_PHASE1.
  SOURCE
  Drosophila melanogaster
  ORGANISM
    Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
    Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
    Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
    1 (bases 1 to 133155)
    Muzny,D.M., Adams,C., Bailey,M., Barberia,J., Blankenburg,K.,
    Bodota,B., Bouck,J., Bowler,S., Brooks,A., Buhay,C., Bunac,C.,
    Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C.,
    David,R., Delgado,O., Deshazo,D., Ding,Y., Domah-Rashid,N.,
    Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D.,
    Forcum-Tansey,J., Frantz,P., Ganesh,R., Gorrell,J.H., Gorrell,L.L.,
    Guevara,W., Harris,K., Hernandez,J., Hodgson,A., Hogues,M.,
    Holloway,C., Hosak,H., Jackson,L.E., Jackson,L., Jia,Y., Jones,M.,
    Kelly,S., Kondejewski,N., Kong,Y., Kovar,C., Leal,B., Li,Z.,
    Lichtarge,O., Liu,J., Liu,W., Logan,O., Lu,J., Lucier,R.,
    Martin,R., Martinez,C., McLeod,M.P., Mei,G., Morgan,M., Morris,S.,
    Nash,S., Nelson,A., Nguyen,R., Nguyen,N., Nguyen,S., Oswal,G.,
    Parish,B., Paxton,S., Payton,B., Perez,L., Pu,L.L., Quiles,M.,
    Reiter,D., Rives,M., Samuel,S., Say,J., Scherer,S., Shah,E.,
    Shen,H., Simon,M., Sparks,A., Stamps,A., Sugang,R., Taber,P.,
    Taylor,T., Vasquez,L., Vinson,R., Vo,Q., Wahbah,M., Watlington,S.,
    Weinstock,G., Weinstock,I.R., Williamson,A., Worley,K., Wren,J.,
    Wrenford,G., Yu,W., Zhou,X., Nelson,D. and Gibbs,R.
  Direct Submission
  Unpublished
  2 (bases 1 to 133155)
  Worley,K.C.
  Direct Submission
  Submitted (16-OCT-1999) Human Genome Sequencing Center, Department
  of Molecular and Human Genetics, Baylor College of Medicine, One
  Baylor Plaza, Houston, TX 77030, USA
  TITLE
  JOURNAL
  AUTHORS
  COMMENT
    * NOTE: This is a 'working draft' sequence. It currently
    * consists of 39 contigs. The true order of the pieces
    * is not known and their order in this sequence record is
    * arbitrary. Gaps between the contigs are represented as
    * runs of N, but the exact sizes of the gaps are unknown.
    * This record will be updated with the finished sequence.
    * as soon as it is available and the accession number will
  
```

```

* be preserved.
  1
  854: contig of 854 bp in length
  1695: contig of 841 bp in length
  2535: contig of 840 bp in length
  3358: contig of 823 bp in length
  4171: contig of 813 bp in length
  5125: contig of 954 bp in length
  6105: contig of 980 bp in length
  6952: contig of 847 bp in length
  7603: contig of 651 bp in length
  8573: contig of 970 bp in length
  10724: contig of 1151 bp in length
  10971: contig of 1247 bp in length
  12519: contig of 1548 bp in length
  14371: contig of 1852 bp in length
  15284: contig of 893 bp in length
  16905: contig of 1641 bp in length
  19536: contig of 2631 bp in length
  21604: contig of 2068 bp in length
  22985: contig of 1381 bp in length
  24383: contig of 1378 bp in length
  26424: contig of 2061 bp in length
  28204: contig of 1780 bp in length
  31215: contig of 3011 bp in length
  33382: contig of 2167 bp in length
  36316: contig of 2934 bp in length
  40441: contig of 4125 bp in length
  44615: contig of 4174 bp in length
  49292: contig of 4677 bp in length
  53647: contig of 4355 bp in length
  59590: contig of 5943 bp in length
  61868: contig of 2278 bp in length
  67152: contig of 5284 bp in length
  71512: contig of 4360 bp in length
  76928: contig of 5416 bp in length
  82257: contig of 5329 bp in length
  89418: contig of 7161 bp in length
  103318: contig of 13900 bp in length
  113206: contig of 9888 bp in length
  133155: contig of 19949 bp in length.

  FEATURES
    Location/Qualifiers
      1. .133155
      /organism="Drosophila melanogaster"
      /db_xref="taxon:7227"
      /clone="RPC198-2017"
  BASE COUNT    38122 a  28987 c  27853 g  38141 t    52 others
  ORIGIN
    Alignment Scores:
    Pred. No.:      5.7e+03      Length:      133155
    Score:          52.00      Matches:      11
    Percent Similarity: 65.22%      Conservative: 4
    Best Local Similarity: 47.83%      Mismatches: 8
    Query Match:    40.31%      Indels:      0
    DB:             2          Gaps:        0

  US-09-847-539a-6_COPY_59_86 (1-28) x AC011910 (1-133155)

  QY    4  LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVallys 23
  Db    99660 CTGGAAGCTCTGTTCAAGCAGGTGGAGTCTCTCCAGCGCCGAGTTCCTCTCTGTGGCA 99719

  QY    24  AlaAspAsn 26
  Db    99720 GCCGACAAT 99728

  RESULT 26
  AC007416
  LOCUS
  DEFINITION
    Drosophila melanogaster, chromosome 2L, region 28B-28B, BAC clone
    BACR08101, complete sequence.
  AC007416
  AC007416
  VERSION
    AC007416.12 GI:14327741
  
```

```
/db_xref="taxon:7227"
/chromosome="2"
/map="28E1"
360
/note="push3420 P element insertion site"
join(370..513,695..789,913..1325,1581..13747,3808..4335,
4404..10389,10447..10651,10701..11393,11455..13118,
13175..13657,13715..15650,15706..17302,17366..17584)
/gene="push"
/product="pushover"
370..17584
/gene="push"
join(395..513,695..789,913..1325,1581..13747,3808..4335,
4404..10389,10447..10651,10701..11393,11455..13118,
13175..13657,15706..15715)
/gene="push"
/codon_start=1
/product="pushover"
/protein_id="AAD20449.1"
/db_xref="GI:4426610"
/translation="MSAHSGGTDWNSVVKALINRTGALKNEVVNLLKATTCRCHDF
FEENSTOFYTAFAAALADKLMOIKTICOTQICQLHDATAVLIRLIYLRPRVSVYE
TKWLLGALKMCEGREGCPASASSMEDYNNAVNLKCKHPSTTKSIMPSSSSGGG
ASNDKSPKSKRSDDLSSVLIQOLIAPLEPKMTWVPLSEETDCTEBOILANVE
YFQONGVDTLDDVCSLPIILNRYRSKYMETINGKSLYLPQTQVEATAVKSNNHML
TDLTILSOAQAIIEMOPLTPSRIERLSKMGIAALYNAVLTISVLGMSOASSQKO
TASTSQSGVGGSGGSKDHDDEQACISIVNKALEIYSNIGMFKTSARIHVYON
HLCYGMILLISIGIOMAGSAGSGSSDSASKSKATKSGSEAGTAPTPIARNLF
KVQOFGELNANAHSKLLSELIEDLVKVEACGQSLESTELPEPAQDILQNYSSL
ERIVVRLNATLHQLFTLATVAYKCACTAKARAKORTECEPISYSDSTYFNDSLS
CSDNSEEDSESYLGHMFKETLSPETHDDNANTSTQERAEQKSAVLPKLDPEPHYLDL
AADIFCDLPQLANRHAYMQRYKAGVSQDQMLLMANIIDKDRDVRNRETDOGSGLA
PASAAGATAGATKQWTSIRFSGAAGRYIHNLISLSELOQSNLLOHLQSLTSP
STDTNTPLOYVSTSLVQIILLLKPTQKEACLSVWHRLINTLVEGVCSSNTASD
SDYEDLNIEHAQALLFELHSLNLMQKKSILLITAGGVIRCAEVCGRGISEDPRVNSOI
MLSRDLLEFLTKMHLNAPPEDLVQVRNLFVSSMPDQKIDLLNCKTKUNSYC
RHDIENFRKSGVSSIRPTFFYSLVGMDPETSVAQEFKDLGIANLILCTPKLK
YPLVDALDTLSLTDMSWYKSKDKKASHNLCAIOYFTIAWNLGLPSTDSHVE
SKAERSPNLHMSIRLPLASSHYLVSSILKQGMVTOYAEIWLTHVGDIGADIKY
SLKOTILGEAFNSOMNGNTPRLSDILFSLVAHQAVAWANKEGLKWPKESEDA
AGEOSAGTSLTSSNDPELYSSNESIDDKLKDDCKLSSDLQKYNQVNLVWKL
NSYQLSVYRGOMKQLSSSTPEKALNIVPVSDRPAIMLEHAAPLKLUPNEDKQ
LIANEWPKLMDNFARNGKHPVEPIYLVNDIAHIIELTSGNSYRSLTLTKHLCKSI
LOFELLLELPHNTANAETVQLQLLISMLDMRTDYLOGHSECLREILSLGTQEAQ
LLLEHVMYGYCYRMKLEFAADLRQAAAGSGGAPLQODRAMFNESMLFVAKTWMFKM
LEKVPVAMROFTKQDORSGLTLLSFTCTSLPVSYARKMLQFVNRFLFOLSLQADS
QOHEDLVECFSELATVDVARLKQWGLHLYXGPNVSTDYSTSEALDTCMLTSLIQP
SSSSSNAQTFTNATVSMSPSISDQDPMIEIYDCGTAAGGSGSNPTGGAAANTSQ
ILSLWQAQPPDSSESSQACDSISERNAGLLLSFVKSVDQSKSOIAPPDFOALL
QLGQTLISPOEGCDFADVLIQIMITLADASPARGHVALFNTLLMLELAKLQLPDKHL
KHAENVSAQLRYLSLQSIGFGRSQHNPPWDDDELQTDIDELYDELAEEGEODSLD
DSEDTLNNKLTCTESQKQEPNQHWHYHCTCNMINTVGCVCARVCHKGHVSYAK
YGNFCDGAKEDSCQALSRRIGSNEVRDSAGTGYLPSHMSLLACKRSLSPVGP
VTRKDSLTNERIAVLTKLEPYRETLOHQDQWLLVYRCILEYFDLLPLSPINENCMXY
SIVGCHRRATAALERLHLLEQSPQTDLMFATLGSQEGAFENVMNITSGOGQTIKH
LLSGVVRARFAERLSPHRRQQLAVSHEKVTIQLSALLKOADAARLKLTLQI
SSAPACTVISAANPCNEDCLAVGKLECHILTFSSSGTNEHIVNVPQENGYIK
KAVPLGQSOTLLAVSDYKIVYDLAVDTYSPKYIYLVAVGKIRDCYFVYODGYNML
TFASSGYLYTQOLDOQSLAVHGVFVNTTLESHQIKDIAHIGGGGVSYIYSHTLQ
LLFYSYSGRSPFFPLTNVSEGVGIYHLDTSASKSKGPIQLQPLQVTEVTHGL
VYASQTSNPIILMITPERIYLOEKAQSAKSRIMDVVGRHRAVAGVETKTLULLCE
DLSURISAQPEPTFVMSLPQVPGNOLYSTLMAGGSGSTSKSKNSTASGKMTS
RKSQOQPPAGQVPPIDFFEHENMLADVEFGNDLLQIYNKQLKTLFTSTGMF
VAPTRNGFTLEVILNNDPNVIVIGVILVITQDVQAPSVTILGRTPTPVRRARWF
DIPTREMLQSKLLKVVFAKAPDPEHVTLLDCEIYVYKSKELVGVDESEDTVPG
SSAPVSSQASSANFEGFNCITQDRMANHLEVMDCALHLLGSGVPASMRQAKV
TASALLPLTPNVTQARYVLYATLYGTRALYHNKDGVLQFVNRRELQSQWPKLEKL
ETUREIDPEAFYVLMVRGIANARPOSIAKICVENNYDIPTLVLELHVKYPTTL
DEPNVIRKGLQCPETIVHCLVEIYMGFALADPGQVGRMTRYFIDLLKHASVISHSA
KEALITLSPMRKRVKVAIVTPPACTPTPTSTSTMOALQAAASAAADIIEAAGVVV
DGSVGGGLPEPNADAEGAAGVGVGVGQOQMLNLEAFNGGCFPRLLGLPEDADEAIM
DIATALSQQHGGDANALQSQGLANLQGIQATMAAAANVLSGGSDDEGS
NVATDGSTLRTSPAEPAGSGSGSGVSGSIGTSARSNSNFGDHANASPPRGSTKD
```

MLLSRLLLFLEYLMKLYNAPPELLDQVRWNLFSVSMPTQKITDILLNCRTKLNSVC  
RHDIEENFKSGACEYSSIRPTFYSILVMGDPETSYWAQEFKLDGLAWNELTCTPDKLK  
YPLLYDALDILSDIMSMYSKEKDEASHNLCATQYCTIANKNLGLPSPSTSHVE  
SLKAERSPNLGMLWSIRUPLASSHYLVSSLLKQGMYYOIAETLHTHVGIDGADIKY  
SLKQTLIYFAFNSQNGTGPLSLDILFDSLVAHQAVAWANKKEGLKRWELKRESEDA  
AGQSAGTSYSSNDPELYSSNESIDDOKKQDDGKSSDLOKYNLVWELMVKIM  
NSYOLLVLVIRGMLKQLSSSTPEKALNLITVPSDKPATMLELHAFLKLLPNEDKO  
LIANEPKCLWDFAFNGKHVEPYILNVIDAHILIELTSGSNYSHTLTKLCLKSI  
LQFPELLPHRTANAETVKQLLSSMLDMRTDYLQHSCHLREILSLGLTEAQK  
LLYEHMGVGYRMLKEFAADLRQPAAGSGGAPLDQDRAMFNESMLFAVLKTFMKM  
LEKPAVQAMRQCFKQDQSGSLTLLTSETGSLPVSYARKMLQFNRLFQLSLQADS  
QFQHDELCFCELSATVVARLKWLGHIIYIGPNVSTGVSTSEALDTCRMLSTILOP  
SSSSNAOTPTNMTVTSAMPSSIDOLDPMEIYDCGTAAGGSGNPGTCGAANSTQ  
ILSLWQAQDPSEESQACDHSERNGALLSFVKSILVKQDSKASQIAPPLFOALL  
QUGTILSPBQECADVLQIMITLADSPARGHVALFNTLLMLELAKLQDPKRLH  
KHAENYASQRLYLSELLQSGIFGRSROHNPDPDELODTIDELYDELAEEGQSDLLD  
DSDEDTLNNKLCFTSQKREFMNRHGYHCHTCNMINTGVCSVCARVCKHGHVSYAK  
YGNFDCGAKGDCGQALSRRIGSVNEVDSAGISGYLPSHMSLLAGKRRSLPVGPQ  
VLTRKDSLNERIAVLTKLLEPYRETLQHDOWLLVVRCTILEYFDLILLPSINENCLY  
SIVGCHKRAFLERLHLLQSPQVTDQLMFATLQSQBGAFFENVRMYSDDQOTIKH  
LLSGVRRVAFCCSLSPHGRQOLAVSHEKGVITILQSALLKQADASKKULITQL  
SSAPIATVLSLAENPCNEDCLAVGLKECHILTFSSGSTEHEIYVNPQLENGNYIK  
KAWLPFGSOTLAVVTSYKYIYDLAVTYSRKYIYLVAVGKIRDCDTFYVDQNGYML  
TFASSGYITQDQOSLAVHGDYVNTLLESHOIKDLAGHGGGSIYIYSHTLQ  
LLFYSYSGRSRFSPTNVSQGVKGIYHLDTSASKSASKGLOPLQVQVTEVTHGCL  
VASMOTSNPILMLTPEIYLOEIKQASAKSRIIMDVGIRHVAAGVETKILLICE  
DGSRLFSQAPETSWLSPQVQFNGLLSILMAKGGSGSTSKSNTASGKMTS  
RKASQOQKPTAGQGVPPIDFHECNMLADVEFGNDLLOIYNKOKLKRFLSTGME  
VASTRNGLETLEVINDPNVIVIGIRVITQDQVAPQSVTILGRITPTVPRARWF  
DIPLTRNEMLQDOKLLKVPKAPADPEHVTLLDCIEVYKSKELWPDSEEDVTPG  
TSAPVSSQASANSFEGNCITOLDMANHLLVMDCALHLLGSGVPAKMQKAVK  
TSALLLLPTNPVQOQARYLATLPTGRALYHNYKDVLLQFVNRLOQSKPKEKL  
ETUREIDPEAFYRLVLMVRIANRQPSLAKICVENNYDIVPTMLVLSELHVKPTPL  
DEPNVIVKRGLOQETIVHCLVEIMYGFALDPQGVGMTKYFIDLLKHDIASISHA  
KEALILLSPKQERRVAVITPACSTPTSTMQALQAASASADSIIEAAGVYV  
DGSVGGGLPEPNADAGAAVGGVQCOQMLNLEAFMGGGPRLLCGLPDDADEATM  
DIALLSQHQDANALQLOGLANLQIGIROATAWAAVANAANAVSLGSDDEGS  
NVATDSTLPTSAPAGSGSGSGSGVESIGTSARSNSFGDHANASPPROGSTKD  
DQOQPSGVAGSGVAVLSNMSSEEDNEANEDKSLKHLDRITLVLESITLQHLGTFD  
LNGLAQIPLITMLTDLNNGERDQVLLHLLTALVGLGKNRKGAAARETRC  
PGNEVRLALLSVGLMGKYKSGTGTTPHQFKDNSFVASTTANVLSKSGAFVYA  
LEALNTLLHVMKVLGDPYAAGGLASQASAGSAGSGVQLLPIKHGPKPDISIL  
IPHNYLKNYPDIIFESYDGLLTIIVPLPYOILRLSSAHPDNYDSGCEAMTFTLCVYM  
MLNLTLLRQVREKLLMYICGSEKRMVDRGHSLOAHFRVVRKVCNIVSSKTCAPYN  
ANPMLSYDALVELTEHLRTQCEISQRTGNWOKFCVYHEDALAMEIACIQLDDGV  
SPIIILLOQAACNLPPPSKQAPQPSSTSSAGKSLRTDREKSEDTDAYSKFPAQ  
CGTFVHOIRYACDALLIREVRIELENITQIRQAHSFMTGLFEHANRGRKELIN  
IFWNLAPVPTYGRTAQFVLLGLYLTLSTRITERLPEVSRADVLRQONELCKH  
PNAPIVTTLESILQVNGYXLESPECLVCNNPEVPMANIKLPVSXSKYTTTMYIKL  
VQCHTISKLVIRLADLKRKMWRTINVIYNNRSVQAVVELKNRPAHKKARSVLSQA  
QTEKIDFPLPITACNLMIEFDFETVSGSENLOQPCRSAAVPAYPGVCGNGENY  
FOCHKRAINIDKDPFLCHSCGFYAKDFDSYARVCCAVDPFIESAEDVRVTSLI  
HSSLERADRNYROLNKNQMLELLIQKVAEHRSDRMVEDNMAVSHSTSOVNRKIQLL  
AOKYCVESRSTFELSKIYOKVACRSELVAYDROQDQPPVNPSTGTAENPTNRC  
YGCALASTEOCLTLIRAMAYNDCRVCLYSGLYSELAENLRGTPLOIEVEYNLLY  
VLTKDNAEACHLLQLVTRVKNALMSIPLISLEAAVHQMTEMLLEVLQGDQDCIMY  
KLKIVFELFISNCLPRGPVTVLHFLCRIMQNLICPVLPSPKNQVATVTDLCMKM  
FEGNTDYKLVNDRNHEYAAWSKRPSNNAKLNKAKDQNAAGSGDAPSKRSRE  
VRVAFYSEKYGKWRERVLQKQVIXPLFVNAKWIOPLLFNANSRQVLOACLSLSL  
SRTNERRQOALNITSLFKHVGAGEASAEYLMYKNMATEQPMLOYLVKLKGVLSQS  
OLLAIREISKVHNMEEVSLSDLSLGLVARYVELLWLLLECPNIRRTYKTRMLGPVLE  
SYIALSLSVQVRLTRIDQAEKLEMLEMDNTSGTEETRAFNEILLDTVEKTRMNDIK  
TVPFVERLYSIITHEPHEDESEFMTLEKDPQOEDFLQGRMLGNPVPSSBMGLPLMR  
DVNKKITCEALLADDNGMELLYNNKIISLDLPVKDYIKVYKRWLAEGEDRAMVR  
YMRGLLOVLAETFEETLNNKSQOVEDTOELRYMANVLADCNKLOMLOSENDSIOANTE  
RNRELQVLLKFLICVKVRCOEVLCOPEIGAINITLLKVLQMLCLOSENDSIOSANTE  
QLEIEMETILSKAASITLDSFLQSLTFPGPEYVSALISCTDCPNRNNRNSVLRHLIR  
VLAAYIETGNEVKNALICEHFKOTLNRFRNDEPTPEEFKLELFCVLTNIOLEHNCIGG  
TKADYIVSUGIVKVALSITETEHAPCVKPTLLRDSDELKEFISRPSLKYTLRFLTGS  
NHREATQVAISKDIIPILHLEQVSSDEHVSLEAENLEALSDSTATAARQVQVRDFT  
RAEKKLANATREKQDLAGMRNTNEKQVTAKSIILOKIEKLDGTCITFCRGYA  
COPDKLVITFTFKCNVVEPEFLKSRKITGYTTVTHFNVVHVDCHTSAIRLTRGRDWM  
ERASLQNAVTRCNGLPLMGPAVGEAASFACMTRHSSYMOESTORCDISYTSVHDLK

LLLVRAWERSPHDAGGGGQSNMHPVPYLLFYSVYVLLSSRSAAARDSKTLTLTYLQA  
PPEKWLCECYEVDGPLFMATISLSLHRSRELNKHKLAHLKRLMVAOGRHIVSPAVLC  
KALLADPADROVKDQTYTKYKPELMMHALVDLIDNLFTKYVSTPKEDWPIISLFDYLRKD  
EALLKSTDSILQTLTBEFLPCTSFVEFCOVAGLHLHLEHPDNFTEELAAALPSTSSN  
\*  
variation 2207 /gene="push"  
/replace="t"  
2672 /gene="push"  
/replace="t"  
5411 .5599 /gene="push"  
/note="putative"  
/bound\_moiety="zinc"  
11402 .11542 /gene="push"  
/note="putative"  
/bound\_moiety="zinc"  
15995 .16135 /gene="push"  
16130 /gene="push"  
/note="push"  
BASE COUNT 4288 a 3943 c 4035 g 3869 t  
ORIGIN  
Alignment Scores: 615 Length: 16135  
Pred. No.: 52.00 Matches: 11  
Percent Similarity: 65.22% Conservative: 4  
Best Local Similarity: 47.83% Mismatches: 8  
Query Match: 40.31% Indels: 0  
DB: 3 Gaps: 0  
US-09-847-539a-6\_COPY\_59\_86 (1-28) x AF096897 (1-16135)  
Qy 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluAlaAlaValVallys 23  
||||| ||||| ||| ::::: ||||| ::::: |||||  
Db 5971 CTGAAGCTCTGTTTCAMGACGAGTGAGTCTCTCCAAGCCGCGAGTGTGCTCTTGTGGCA 5912  
Qy 24 AlaAspAsn 26  
||||| |||||  
Db 5911 GCCGACCAAT 5903  
RESULT 23  
AF096896/c 17595 bp DNA linear INV 15-OCT-2001  
LOCUS Drosophila melanogaster pushover (push) gene, complete cds.  
DEFINITION  
ACCESSION AF096896  
VERSION AF096896.1 GI:4426608  
KEYWORDS fruit fly.  
SOURCE  
ORGANISM Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
REFERENCE 1 (bases 1 to 17595)  
Yager,J., Richards,S., Hekmat-Scafe,D.S., Hurd,D.D., Sundaresan,V.,  
Caprette,D.R., Saxton,W.M., Carlson,J.R. and Stern,M.  
CONTROL OF DROSOPHILA PERINEURIAL GLIAL GROWTH BY INTERACTING  
neurotransmitter-mediated signaling pathways  
Proc. Natl. Acad. Sci. U.S.A. 98 (18), 10445-10450 (2001)  
PUBMED 11517334  
REFERENCE 2 (bases 1 to 17595)  
Richards,S., Hekmat-Scafe,D.S., Hurd,D.D., Caprette,D.R.,  
Saxton,W.M., Carlson,J.R. and Stern,M.J.  
AUTHORS  
DIRECT SUBMISSION  
SUBMITTED (06-OCT-1998) Biochemistry and Cell Biology, Mail Stop  
140, William Marsh Rice University, 6100 South Main, Houston, TX  
77005, USA  
FEATURES  
source  
1..17595  
/organism="Drosophila melanogaster"

MLFAVLTKMFKLEKPVAVQAMRQFFKQDRSGSLTLLLSFTGTSLSVSVARMLQV  
NRLQSLQADSOQHEDLVECESELAIVDVARLKWGLHIIYGPVNVSTVSEALD  
TTCMLSTSIQPSLSSNAQTNTNATVSAMPSISDQDLPMEIYDCGTAATVSEALD  
PGTGAANTQILSLMQAAQPNPSESSQACSDSERGALLSVKSLVSKDQSA  
SOIAPQALLOLQOTLIGPPOEGCDFADVLOMITLADASPARGHVALFNTLLWL  
ELAKLOLPDKHLKHAENVSAQLRVLSLQSIGRSGROHNPWDELOYDIDELDE  
LAERGQDLSLSDDEDTLANKICTFSQTKQEPNQRWYHCHFCNMINTVGVSCVAR  
VCHGHDVSTAKYGNPCDCGAKEDSGQALSRKISGNEVRSAGISYLPUSHSLIA  
GKRKSLQOPVLFTNDRSLTNERIAVTLKLEPYRETLDQOMLQVCEILEYFDL  
LLPSINENCMYLSVTCGRKTRALERLHLEQSFQVTDQMFATLGSQCEAFVNRW  
NYSQDQOTIKHLLSGVVRVAFCLSSPHGRQOLAVSHEKGVTLIOLSALLKOA  
DASKRKLTLQLSAPACTVLSLAANPCNEDCLAVGLKECHLITSSSGSNEHIV  
VNOLENGTIIKRAVLPQSGTSLIAVVTSDYKIYDIADVTISPKYIYLVAVGRDC  
TFVYQDYNMLTFASSGYITQOLDOOLAVHGFYVNTLLESHQIKIDAGHIGG  
GGVSYIYSHTLQLLFYSGRSPFSLTNVSEGVGIYHLDTNASAKSKGPIQL  
VOMTEVGHPLVYVAMQTSNNPILMITPERIYLOIKQAQSAKSRIMDVVGRHVA  
GVEKTLILLCEGSLRIESAOPETTSFWSLPOVQFPGNOLYSTLMAKGGSGTSK  
SKSNTASKMTSKASQOQKOPTAGGQVPEIDPEHCNMLADVEFGNLLQIYNKQ  
KLKTRLFSTGHFVASTNSGFTLEVINNDPNVIVGRVIGTODVQRAQSVITLGR  
TIPTVRARAFDILPTREEMLOSKLLKVVFAKAPDEPHVTLLDCEVYKSKELVY  
WPDESDTVPGSSAPAVSSQASANFEGFCNITQIDRMANHLLBVMDCALHLLGS  
GVPSMRKAVKATASALLPTPNPQOTQARYLATITGTRALYHNTKGVLLQFVNR  
ELOSMPKLEKLETLREIDPEAFVRLVMVRGIANARPOSIAKICVENNDIVPTLMG  
IVLEHVTPTLDEPNVIVKRGLOPETIVHCLVEIMYGFALADPGQVGMKTFIDL  
LKHDASVSHSAKALLILLSPRMKRRVAIVTPACSTPTSTSTQWALQAAASSA  
SDIIEAAGVVDGSGGGLPEPNADAEGAAGVGVGOQOMLNEAFMGGPPRL  
GLPEDADEIMDIAIALSLQGHGDANALQSPQOGLANLQIRIATMAAANVAAAN  
VSLGSDDEGNSVATDGTSLRTSPAPAGSGSGSGVIGETSAKSNFGDHA  
NASPPROGTDKDOQPGSPGAVGAVLSAMSSSEDNEANEDDKLSKHLDIRAV  
LESIIHLGTFDLGNGLOAPIQVILMLTDLGNRNOQOVLHLLTALVIVEIG  
KVAASARMEKTCNEVRLALLSLFGVMGKTSKQGTGTSPPHOFKDNDSFVASTTA  
NRUSKSAFYALPANTLILVHMKNVLGDPYAGAGLASQASQASGAGSGVOLLKQ  
IKHGPFDISILPHNTKNYPDIFESYDGLLTIIVRLPYQILRLSSAHPDNVDYFC  
CEAMFTPLCYMLNLTLLRQVRKLIMICGSKERFMRVDRGHSADHVFVVRKY  
NIVSSKTAGVNPMLSYDALVELTLEHTCOEISQMRGTGNQKFCVVEDALMI  
MEIACYOLDGVSPIILQLQAAVCLNPPPSGSKQAOQPSSTSSAKLTDREKSD  
TDAYSFDPACQCTVHQIFRYACDALIRFRIFLENNITOLRWAKSDFMTGFE  
HANERQELLKFNWNLVPTTYGRTAQVDLGLYTLTSTRITELPRLFSRAYD  
VLRQNELLKHPNAPTYTLESILQVNGVYSEPCLVNCPNPEVPMANIKLPSKVS  
SKYTTMTYIKVOCHTISKILVIRADIKTKMVRTINVYNNRVSQAVLKNRPL  
WHKARSVLSAQTELKIDPLPITACNLMTETAPDEFYVSGSSENCQPCRSAAVA  
YPGVCGNGENVFOCHKRAINDEKDFELCHSGCFCKYAKFDFSMYARVCCADPIE  
SAEDRVYLSLHSSLRADNRQLTNKMOLELLQKVAHRSSDRMVEDNNASVH  
STSOVNKIIQLLQKYCESKTSPEELSKIVQVKACRSELVAYDROQOQPNVPS  
TTGAENPTNRYCALASTEQCLTLRAMAYNDYCRVCLYSQGLSELAEHPNRRGT  
PLIQEEVNRNLIVLTKNACAPMHLLOLTVTRKNALMGSLPLISLEAAVHOENTLE  
VLIGQDDICWEYKLVIFELFISNCLPGRPVAVLHCLRIMONLIPCVLPGSKPNQ  
KVATDLCMKMFCGTVDYRAWLNDRNHEYAANKRMPNSNQAKLNKADQNVAS  
GGSADPKSREVRVAFSEKIGARWRVLDKQVIKPLVFNAKIOPILLFNANSRP  
GRQACSLLSLSTREROOALNMLTSFKHVGGEAGEASAEYLMYKNMATEQPLQ  
YLVKGLVLSQSLLAIEISKVHMEYSLSLSDLSGLYALQYVELLWLLCEPNIR  
TYKTRMLGPVLESYALRSLVQRTLIDDAQEKLEMDTSGTEETRAFMELLIT  
DTVEKTRMNDIKTVPVFERLYSIIHPEEHDESEFYMTLEKDPQOEDFLOGRMLGNY  
PSEMGKGLPMDRVNKNICTDCELIALLEDDNGMELLYNNKIIISLDLPKDYKKVWL  
AEGDGRAMIRVYMRGLGLGDTEFEVETLNNKSGEQVDTEQYRMANVLADCNGLV  
MLERIGLSLRNRELIQVLLKFLICVKVRRCQVCEIQPEIGAINLLKVLQMCLO  
SENDSIOSARHETLEMETILSKAASDCEHFDLSLTFEGGVEYVALISCTCPNV  
RNPSPVLRHIVLAAVLVYGVNEKMAIICFDKFTLNFNRDNERTPEEFKLEFV  
LTNOIENHCIGTKDYIVSLGIGERSLAYTEHAPCVKPTLLRTDSDDELKEFTSRPS  
LKVITLFLTGLSNHEATOVAKIDIIPIHRLQVSDHVEHGLSLENLLEALSDSA  
TAARVQVQDPTREAKRLAMATREKQDLDALCMRTNEKQVATKGSILQTEKURDET  
GLATCFICREGVAFQVGLVITFTKCNVEEFELKSRKTIYTTVTHSVHVDYDCT  
SAILTRGDEWERSAQONATRCNGLPLWPGVAGEAFAFSACHTSHVNMQESTQRC  
DSTYTLTYLOAPPSEKWLCEGYEVDGAGGPGQSNMHFVPLLYFVSYLLLSRSAA  
ROKRLTYLQAPPSEKWLCEGYEVDGAGGPGQSNMHFVPLLYFVSYLLLSRSAA  
QGRHVSAPVLCALLAPADQVADTVYKPLFMALWALVDLIYDNLKFTYSTPREDM  
PISLFDLIRKNDKALLSTUSILQTLTEEFPLCTFSYVEFCVAGLLHLHLEPQNFIE  
ILAAPSTSSSN\*

BASE COUNT 3257 a 3028 c 3148 g 2940 t  
ORIGIN

Alignment Scores:  
Pred. No.:

465 Length: 12373

Score: 52.00 Matches: 11  
Percent Similarity: 65.22% Conservative: 4  
Best Local Similarity: 47.83% Mismatches: 8  
Query Match: 40.31% Indels: 0  
DB: 3 Gaps: 0

US-09-847-539a-6\_COPY\_59\_86 (1-28) x DME17920 (1-12373)

Qy 4 LeuGluAlaLeuAlaAspGlnThraspAlaLeuGlnSerGluGluAlaAlaValVallys 23  
|||||  
Db 2329 CTGGAAGCTGCTGTTCAACAGGTGGAGTCTCTCCAAAGCCGACGAGTCTCTCTTGTGGCA 2270  
|||||

Qy 24 AlaAspAsn 26  
|||||

Db 2269 GCCGACAA 2261  
|||||

RESULT 22  
AF096897/c 16135 bp mRNA linear INV 15-OCT-2001

LOCUS AF096897 Drosophila melanogaster pushover (push) mRNA, complete cds.  
DEFINITION AF096897  
ACCESSION AF096897  
VERSION AF096897.1 GI:426609  
KEYWORDS fruit fly.  
SOURCE Drosophila melanogaster  
ORGANISM Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
REFERENCE 1 (bases 1 to 16135)  
AUTHORS Yager,J., Richards,S., Hekmat-Scafe,D.S., Hurd,D.D., Sundaresan,V.,  
Caprette,D.R., Saxton,W.M., Carlson,J.R. and Stern,M.  
TITLE Control of Drosophila perineurial glial growth by interacting  
neurotransmitter-mediated signaling pathways  
Proc. Natl. Acad. Sci. U.S.A. 98 (18), 10445-10450 (2001)  
JOURNAL 11517334  
PUBMED  
AUTHORS 2 (bases 1 to 16135)  
REFERENCE Richards,S., Hekmat-Scafe,D.S., Hurd,D.D., Caprette,D.R.,  
Saxton,W.M., Carlson,J.R. and Stern,M.J.  
TITLE Direct Submission  
JOURNAL Submitted (06-OCT-1998) Biochemistry and Cell Biology, Mail stop  
140, William Marsh Rice University, 6100 South Main, Houston, TX  
77005, USA

FEATURES  
source  
1..16135  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/chromosome="2"  
/map="28E1"  
<1..16135  
/gene="push"  
<1..25  
/gene="push"  
26..15994  
/gene="push"  
/codon\_start=1  
/product="pushover"  
/protein\_id="AAD20450.1"  
/db\_xref="GI:426611"  
/translation="MSAHSQGTWNVSVKALILNRTGALKNEVVNLLKATTCRCHDF  
FEESNTQFTAFALAAADKLMQIKTICQIQQLHDATVLIIRFIYLRPVSYVE  
TKWLGLMLCEGRECPASASSNFDINAVNKLKCKHPSTITKSPSSSSSGSG  
ASNDKEPKSIKRSRDLSSVLIQLIAPLEPGKMTWVPLEETDCTEGLAANVE  
YFQNGQVTLDDVCVSLPILNRYSKYMETINGKSLYLPLOVEATVAVSMNHML  
TDLIISQAOALIEWOPILPSRIENSCGTAALYNAVLTISVLCMSAASSQKQ  
TASVSGSGVSGSGGOSKDDHDFEQACISVNAKLEIYNIGHMFKTSARIVYON  
HLCVGNLLISGICGACSGSGSSDSASKASKATKSGSEAGTAPTTIARVNF  
KVOQGFELNATIANHSITKLSLIEDLKVCAACGQSLSEPLEAQFDILQNYSSL  
EYRIVLVNTATLHQLPFTFLATVYAKCTLKRAKDRTECEPISYSDFTTFNDSL  
CSNSEDSDSEYLGHWFKETLSPETHDDNANTQERAEOKSALVPLKDEPHFYLDL  
AADTFCFLDQFLANRHAYMORYKAGVSDQOMLLMANIKDFDRVMRNETDQSGNA  
PAASAGAGTSAGTSKQTSMLRFSGAAGRYIHNLISTSLSEQLSNLLOHLSLSPW  
STDYNTNPLOVYPTSLVLIQILLLKPTQKEACLSVWHRLINTLVGEGCSSNTASD  
SDYEDNLNTEHAQLLLFLFHSUNLMOKKSIILLTATAGGVIRCAEVCRGISEDPRVNSQI

```

/db_xref="GI:9948367"
/translation="MNSNPLVDWLLDSLEVDTSIFPHVGRYCCGWHASTHGLARASFH
LVVQHCWAHVAGSAGWRGLSDGDALEFRLDVFELSSSEATAEAARCPQRMQALEW
NAADQAGVCCVGFHFHSGLSGLIVDALPDLIRAGDPAADARLFAIIVDECRQP
APSPVLLERLHLLEYLVROQVLDNPTLGGVALARQPOFAGLLEALTEAPEQWSL
EQMARCTGLSRSAFFKRENLGSGQSPGVLLALRMHRACALLRADASVAEVAAGVYQ
STAAFTRAFHAKTGOQPAYRKASR"
complement(5800..7416)
/gene="PA2333"
complement(5800..7416)
/gene="PA2333"
/codon_start=1
/transl_table=11
/product="probable sulfatase"
/db_xref="GI:9948368"
/translation="MSQPKPIRNVLYIMCDQLRRDYLSCYGHPYILHTPNIDRLAAEGV
RPARAYTOCTTCGPRSMSSYTRGVYSSHQVAMNAVPLPLEELTLGDYLRACGVRTALV
GKTHATNLEGMRRIQIDPASARGAALAEAGPEYDRNDGVYDDPAADKREAPYT
HYLRBLGTGONPHDWMANAAGAGEILSGWRMHAGLPTRLPEAHSETATYTRAM
DFIDQGERPCLLUSYIKPHWYITAPYIHALYRADQVLPALRAAPGEESDHPYIRA
FHEHSLNFSEDVRRQVYPTIYMGILIRQVDDQLGRLFOHMRASGRWDDTLIVFTSDH
GDFLGHGLGEKESLFPFVHGAPPDMRRYAIAYDYAFQAPARERLGRPIDACRMVY
AVEPASHYIEGRSLPFGVAGPDMRRYAIAYDYAFQAPARERLGRPIDACRMVY
RSEWKYIAYDGFRAQLFDLASDPCGLRDLGADPAHAAREAHACHLFDLWGLKRR
TISNAEIDLRCQAFRYGPEGRVLPGLW"
complement(7513..8424)
/gene="PA2334"
complement(7513..8424)
/gene="PA2334"
/codon_start=1
/transl_table=11
/product="probable transcriptional regulator"
/db_xref="GI:9948369"
/translation="MHIDLRLRHFIALHEHRSFVAALAVNLQSASFSSRSTOALEHS
ACRLVDRAAKDLRPTKQGLIVLEHARLVSCAHNLAEFLNGLGELRFGAGPA
PATWLVPRAVGVENVYPKARVFOYEHQALNRRLQAEFEFFVADRHFRFADPQYR
TSRLRPRWHCCRGHPLAARSAEELLSFPLATNIPNIRKVLVDLSGRDPDR
PAICENGLSLGVVLAISDAIGISSGLGLDAPPGGGLRPLRVRELDDLDLHTRYG
IVSRAGYSLSLPLAEAMIDELRVDDAT"
8574..10943
/gene="PA2335"
8574..10943
/gene="PA2335"
/codon_start=1
/transl_table=11
/product="probable TonB-dependent receptor"
/db_xref="GI:9948370"
/translation="MNRPAVRSIFSSRAPPFRRLPLSLICACAGCAQLAGADDT
AGTEARQAQPAITLQTVTVTSRRRVENSQDVTAMSVLGGETLESQRIYRVODLQ
VPTWVYVHARQSVSRGLGNPASDGLSGVLYLDNVYLRPGMASFDLLDIEQ
LEVLRGPQCTFGKNTTAGVINIGTRAPSTPERSVEISGGEDGYFOAGRTISGLPE
TLAARLSAYRTVDDCYVKNLHDGNTLNGCAROGFRGOLLFKPNDPFLSLRWIGDYNEE
SDNGVLTFFSSGPTVGVNRYEQRAAAGATLVDCGRKRVNFDSDQVTVFQGGTSVE
ANWLANDFTLTSISAYRWKDFTPRNDGLDVPAYNAGSVVRDKQWSEIRLASPSG
GFFDVLGAYFGQDLDNKSFYDGPADIMNGTPAGALSDVSTRGRHIDVSYALF
AOGTWHLTRDLTAVRGVTEESAWYERAPLGGVAVSAAAAAANGRVAGYSDG
IGLHVSFSAIISLYRFSDOVLGYASLSHGKSGVNLTVASAPAGADSLLIGTER
ANDAEGLKTLTLDNRLTLNGLFWTVHGYOANAYDEANRTQYLTNAGDVSRLGF
EATALPIRGLTLNFNASYNDVRYLSYKNAPCAPEVAFQAGAPASCDLSGHQVVGASKY
IANFNGEYRWKLDGDLLEPLTASYAFRSRAGVTIDDSAYGQIPSYALNAGSLRGDY
GDQVNDLSLKNLKNFADKTYFTSLMNSANGVAGVLGTPRTLGVTRYDF"
10966..12306
/gene="PA2336"
10966..12306
/gene="PA2336"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/db_xref="GI:9948371"

```

gene

CDS

gene

CDS

gene

CDS

gene

CDS

```

/translation="MPRFNMLLSAALLAAGAAHAPSVYPTGVTRYPDOKAFNQYVIF
SGADRKTHLIDMNGNEVKRWDPGPPAAVIDPALNCKKRGHVLLQLADLAADPKLASP
GNGLANQALGELDWDGKEVYMRWGDQAPGAHQHDDRLRANGNTLLLANKVHIAVKG
AYPRVTDVYIYEVTPQGLAWSATSEHLDEFGFTADQLKRVRASENPDYLIHNNLAP
LGENRWAAGDRRFAPDNLLDSRNANFTATVDKASKGVVRLGPNLPLIDPKSARKL
PRPVDIISQDHAIPEGLPGAGNLLVFDNQDAGYPSVPRGLVSGSRVLEIDPLKK

```

Alignment Scores:

Pred. No.:	388	Length:	12411
Score:	52.50	Matches:	14
Percent Similarity:	71.43%	Conservative:	6
Best Local Similarity:	50.00%	Mismatches:	3
Query Match:	40.70%	Indels:	5
DB:	1	Gaps:	1

US-09-847-539A-6\_COPY\_59\_86 (1-28) x AE004659 (1-12411)

Qy 5 GluAlaLeuAlaAsp-----GloThrAspAlaLeuGlnSerGluGluAla 19

Db 10151 GAAGCGGTAGCTAGCGCTGAGCAGAGCGGAGCGGACCTGTGTCAGGCCGATGTCC 10092

Qy 20 AlaValValLysAlaAspAsnAla 27

Db 10091 GCTGTCTAGCGCGCCAGCCCTGCC 10068

RESULT 21

DMEI7920/c

LOCUS

DEFINITION Drosophila melanogaster mRNA for calossin protein.

ACCESSION Y17920

VERSION Y17920.1

KEYWORDS calo gene.

SOURCE fruit fly.

ORGANISM

REFERENCE

AUTHORS

TITLE

MEDLINE

REFERENCE

AUTHORS

JOURNAL

TITLE

JOURNAL

FEATURES

source

source

gene

CDS

DMEI7920 12373 bp mRNA linear INV 18-NOV-1998  
Drosophila melanogaster mRNA for calossin protein.  
Y17920  
Y17920.1 GI:3893108  
calo gene.  
fruit fly.  
Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 12373)  
Xu, X.Z., Wes, P.D., Chen, H., Li, H.S., Yu, M., Morgan, S., Liu, Y. and  
Montell, C.  
Retinal targets for calmodulin include proteins implicated in  
synaptic transmission  
J. Biol. Chem. 273 (47), 31297-31307 (1998)  
99030403  
Montell, C.  
2 (bases 1 to 12373)  
Direct Submission  
Submitted (11-AUG-1998) C. Montell, The Johns Hopkins University  
School of Medicine, Department of Biological Chemistry, 725 N.  
Wolfe Street, Baltimore, MD 21205-2185, USA

Location/Qualifiers  
1..12373  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"

1..3316  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/map="28E1-2"  
/tissue\_type="retinal"  
/dev\_stage="adult"

2..12352

/gene="calo"

/gene="calo"

/note="calossin"

/codon\_start=1

/product="CALO protein"

/protein\_id="CAA76940.1"

/db\_xref="GI:3893109"

/translation="YQLISVIVRGQMLKOLSSSTPEKALNLIPTVSDKPAIMLELHA

AFLKLLPNEDKOLIANEWPCLVWVNDFAFGKHQVPEYILNVIAHIIELTRGSNYS

TLHTLKHCLKSLIQLPELLPHRTANAETOLKOLLISSMLDMRTDYLQGHSEHCLR

EILSLGTQEAQKLLLYEHMVGYCYRMLKEFAADLROPOSGAGSGGAPLDQDRAMPNES



misc\_feature /note="assembly\_name:Contig51"  
120563..130242  
misc\_feature /note="assembly\_name:Contig52"  
130343..140588  
misc\_feature /note="assembly\_name:Contig53"  
140689..149261  
misc\_feature /note="assembly\_name:Contig54"  
149362..162349

Alignment Scores:  
Pred. No.: 5.23e+03 Length: 173772  
Score: 53.00 Matches: 11  
Percent Similarity: 77.78% Conservative: 3  
Best Local Similarity: 61.11% Mismatches: 4  
Query Watch: 41.09% Indels: 0  
DB: 2 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x AC087862 (1-173772)

Oy 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaVal 21

Db 173045 CTGGAGGCTCTGCTGGACAGACAGACGGGCGACAGCAAGTAGCCTTG 172992

RESULT 20  
AE004659/c  
LOCUS AE004659 12411 bp DNA linear BCT 30-AUG-2000  
DEFINITION Pseudomonas aeruginosa PA01, section 220 of 529 of the complete genome.

ACCESSION AE004659  
VERSION AE004659.1 GI:9948361

KEYWORDS Pseudomonas aeruginosa.

SOURCE Pseudomonas aeruginosa

ORGANISM Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;

Pseudomonas.

REFERENCE 1 (bases 1 to 12411)

1 Stover, C.K., Pham, X.-Q., Erwin, A.L., Mizoguchi, S.D., Warren, P., Hickey, M.J., Brinkman, F.S., Hufnagle, W.O., Kowalik, D.J., Lagrou, M., Garber, R.L., Goltry, L., Tolentino, E., Westbrock-Wadman, S., Yuan, Y., Brody, L.L., Coulter, S.N., Folger, K.R., Kas, A., Larbig, K., Lim, R., Smith, K., Spencer, D., Wong, G.K., Wu, Z., and Paulsen, I.T.  
Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen

Nature 406 (6799), 959-964 (2000)  
20437337

REFERENCE 2 (bases 1 to 12411)

1 Stover, C.K., Pham, X.-Q., Erwin, A.L., Mizoguchi, S.D., Warren, P., Hickey, M.J., Brinkman, F.S., Hufnagle, W.O., Kowalik, D.J., Lagrou, M., Garber, R.L., Goltry, L., Tolentino, E., Westbrock-Wadman, S., Yuan, Y., Brody, L.L., Coulter, S.N., Folger, K.R., Kas, A., Larbig, K., Lim, R.M., Smith, K.A., Spencer, D.H., Wong, G.K., Wu, Z., Paulsen, I.T., Reizer, J., Sailer, M.H., Hancock, R.E.W., Lory, S., and Olson, M.V.  
Direct Submission

Submitted (16-MAY-2000) Department of Medicine and Genetics,  
University of Washington Genome Center, University of Washington,  
Box 352145, Seattle, WA 98195, USA  
Location/Qualifiers

source 1. 12411  
/organism="Pseudomonas aeruginosa"  
/strain="PA01"

/db\_xref="taxon:287"

complement(353..1117)

/gene="PA2327"

complement(353..1117)

/gene="PA2327"

/codon\_start=1

/transl\_table=11

/product="probable permealase of ABC transporter"

/protein\_id="AAG05715.1"

/db\_xref="GI:9948362"

/translation="MRQTLRGALGVAGLTLFALWLVGVALFGAAGDLSARSPQAT

LLSLGELLERRELYEHVLSLKRVLGLFALLVGVPLLVGASRNLEAATTPAFQF

LRMISPLSWMPIAVMLMGVDQPIYFLLAFAPALWPIILLNTAAGVQRLDPRWLQLSRSL  
AATPWETLRVILPGILGLVLTGTVRIGAIGLWLVPCPEMLGVSAGLGYILDTDRSL  
AYSELMAVLILGLLGLDAAARGLYRRWAHGAIG"

gene

CDS

complement(1114..2313)

/gene="PA2328"

complement(1114..2313)

/gene="PA2328"

/codon\_start=1

/transl\_table=11

/product="hypothetical protein"

/protein\_id="AAG05716.1"

/db\_xref="GI:9948363"

/translation="MCLDDPTHSSRRDILKLAALLSAAGALPLSLSSLQARAABPDAPV  
RIGYLPITDAPLTVLVAHANGHFAEGIQAEPRVLLRSWAQVIEAFISQVNVHLLSP  
MTVMARYGSKVPKAVVAVNHVGGSLTVAPEDVQRLGGKVAIPFWYSIHNVLQO  
LLRDLGLRAVSRAVGAALAADEVNLVLPSPDMPALASKRIGHYIVAEFNRALAEAL  
KVGVRQFTGDMVRNHACCVVHEHDLRRPOMSKVNAIVKAQLWREHRAEAAQ  
LLSKAGANRYTPHAPENVGLAPGAEEQQAYLASGAIRHADWQERIDIQQYPPYSY  
TEELVRLKDTLIEGNNFASLASDPROVASDLVDVFRVQLAEVGGVLAQFGLAEGFQ  
RSEIEISp"

gene

CDS

complement(2279..3115)

/gene="PA2329"

complement(2279..3115)

/gene="PA2329"

/codon\_start=1

/transl\_table=11

/product="probable ATP-binding component of ABC transporter"

/protein\_id="AAG05717.1"

/db\_xref="GI:9948364"

/translation="MSEALLDARGIALGYPRGAGQAVLGDFFELRLGAGEVVSILGPS  
GVGKSLRLVLAGLQEPAGSVRLGPHGVAFAVQDFPSLLPMLNLEKNVAFG  
LDFARQPHLDAATQRAVDQAIATAVGLHEHARARYPAOLSGGMAORTALARCLEARQV  
LLDEPFGALDEVTRADMQQLLRATHERGTVVLITHDIDEALLSERILLGLDSPA  
RTLGEWRIDLQPRAEELVEELGALRIELTLRASRTHAHPITQAPEASHVPGRPDPF  
PT"

gene

CDS

complement(3112..4179)

/gene="PA2330"

complement(3112..4179)

/gene="PA2330"

/codon\_start=1

/transl\_table=11

/product="hypothetical protein"

/protein\_id="AAG05718.1"

/db\_xref="GI:9948365"

/translation="MLDPTLSDWLDROADLDRGDADPAALLPLLAAGVLGVAVER  
LGAGGTGLADAVERYAAVARSLLTAFFVWQORAFIEYLQSPNQLRORLLGDLLAG  
RLGATGLSNMRFSLGTEALQVVAHAEAGWRLDGLRPVYNLRSGDFVVAALIEHG  
EGKPFVLAIPGELQORSADLRLGLQCSNTAALDRVAVAGRDWLLHDDARQELP  
RVRFATGLQCGMAIGLARALDEVORHLGSGRSLLDGEAAQRETLDGHVAALHAGL  
ASGEFASQPARLFRITRIGLAFAAANAVALQASGGKAYLSEHGAGFARRWRESAFVP  
IVTPSLVOLRAELQROAEAKA"

gene

CDS

complement(4229..4789)

/gene="PA2331"

complement(4229..4789)

/gene="PA2331"

/codon\_start=1

/transl\_table=11

/product="hypothetical protein"

/protein\_id="AAG05719.1"

/db\_xref="GI:9948366"

/translation="MSRVPVLTLENAPAPRPFLQTALNNSGYIPNLGLVLANAAPAL  
ETYLTVSLNAKASLGLPEREVQVLAATHGCDPCVAGHTAVARNKARLPPEVPEAL  
RARGELPDARYEALADPFAVIAIRSGAVTDEQFEAFRAAGFDDQAQALEVILGVSATL  
CNFANVTARTPLNAELFQYRWTFNQA"

gene

CDS

4894..5799

/gene="PA2332"

4894..5799

/gene="PA2332"

/codon\_start=1

/transl\_table=11

/product="probable transcriptional regulator"

/protein\_id="AAG05720.1"

\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 1282: contig of 1282 bp in length  
\* 1283 1382: gap of unknown length  
\* 1383 2426: contig of 1044 bp in length  
\* 2427 2526: gap of unknown length  
\* 2527 3572: contig of 1046 bp in length  
\* 3573 3672: gap of unknown length  
\* 3673 6159: contig of 2487 bp in length  
\* 6160 6259: gap of unknown length  
\* 6260 8183: contig of 1924 bp in length  
\* 8184 8283: gap of unknown length  
\* 8284 11014: contig of 2731 bp in length  
\* 11015 11114: gap of unknown length  
\* 11115 12517: contig of 1403 bp in length  
\* 12518 12617: gap of unknown length  
\* 12618 14315: contig of 1698 bp in length  
\* 14316 14415: gap of unknown length  
\* 14416 17095: contig of 2680 bp in length  
\* 17096 17195: gap of unknown length  
\* 17196 19263: contig of 2068 bp in length  
\* 19264 19363: gap of unknown length  
\* 19364 22746: contig of 3383 bp in length  
\* 22747 22846: gap of unknown length  
\* 22847 24580: contig of 1734 bp in length  
\* 24581 24680: gap of unknown length  
\* 24681 27136: contig of 2456 bp in length  
\* 27137 27236: gap of unknown length  
\* 27237 30151: contig of 2915 bp in length  
\* 30152 30251: gap of unknown length  
\* 30252 33090: contig of 2839 bp in length  
\* 33091 33190: gap of unknown length  
\* 33191 36912: contig of 3722 bp in length  
\* 36913 37012: gap of unknown length  
\* 37013 40616: contig of 3604 bp in length  
\* 40617 40716: gap of unknown length  
\* 40717 44847: contig of 4131 bp in length  
\* 44848 49447: gap of unknown length  
\* 49448 50238: contig of 5291 bp in length  
\* 50239 50338: gap of unknown length  
\* 50339 53196: contig of 2858 bp in length  
\* 53197 53296: gap of unknown length  
\* 53297 57538: contig of 4242 bp in length  
\* 57539 57638: gap of unknown length  
\* 57639 62554: contig of 4916 bp in length  
\* 62555 62654: gap of unknown length  
\* 62655 66368: contig of 3714 bp in length  
\* 66369 66468: gap of unknown length  
\* 66469 71547: contig of 5179 bp in length  
\* 71548 71747: gap of unknown length  
\* 71748 74947: contig of 3200 bp in length  
\* 74948 75047: gap of unknown length  
\* 75048 79238: contig of 4191 bp in length  
\* 79239 79338: gap of unknown length  
\* 79339 84833: contig of 5495 bp in length  
\* 84834 84933: gap of unknown length  
\* 84934 89804: contig of 4871 bp in length  
\* 89805 95932: gap of unknown length  
\* 95933 95932: contig of 6028 bp in length  
\* 95933 102531: contig of 6499 bp in length  
\* 102532 102631: gap of unknown length  
\* 102632 111305: contig of 8674 bp in length  
\* 111306 111405: gap of unknown length  
\* 111406 120462: contig of 9057 bp in length  
\* 120463 120562: gap of unknown length  
\* 120563 130242: contig of 9680 bp in length  
\* 130243 130342: gap of unknown length  
\* 130343 140588: contig of 10246 bp in length  
\* 140589 140689: gap of unknown length  
\* 140690 149261: contig of 8573 bp in length

\* 149262 149361: gap of unknown length  
\* 149362 162349: contig of 12988 bp in length  
\* 162350 162449: gap of unknown length  
\* 162450 173772: contig of 11323 bp in length.

## FEATURES

source  
1. .173772  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="RP11-724H13"  
1. .1282  
/note="assembly\_name:Contig17"  
1383. .2426  
/note="assembly\_name:Contig20"  
2527. .3572  
/note="assembly\_name:Contig22"  
3673. .6159  
/note="assembly\_name:Contig23"  
6260. .8183  
/note="assembly\_name:Contig24"  
8284. .11014  
/note="assembly\_name:Contig25"  
11115. .12517  
/note="assembly\_name:Contig26"  
12618. .14315  
/note="assembly\_name:Contig27"  
14416. .17095  
/note="assembly\_name:Contig28"  
17196. .19263  
/note="assembly\_name:Contig29"  
19364. .22746  
/note="assembly\_name:Contig30"  
22847. .24580  
/note="assembly\_name:Contig31"  
24681. .27136  
/note="assembly\_name:Contig32"  
27237. .30151  
/note="assembly\_name:Contig33"  
30252. .33090  
/note="assembly\_name:Contig34"  
33191. .36912  
/note="assembly\_name:Contig35"  
37013. .40616  
/note="assembly\_name:Contig36"  
40717. .44847  
/note="assembly\_name:Contig37"  
44948. .50238  
/note="assembly\_name:Contig38"  
50339. .53196  
/note="assembly\_name:Contig39"  
53297. .57538  
/note="assembly\_name:Contig40"  
57639. .62554  
/note="assembly\_name:Contig41"  
62655. .66368  
/note="assembly\_name:Contig42"  
66469. .71647  
/note="assembly\_name:Contig43"  
71748. .74947  
/note="assembly\_name:Contig44"  
75048. .79238  
/note="assembly\_name:Contig45"  
79339. .84833  
/note="assembly\_name:Contig46"  
84934. .89804  
/note="assembly\_name:Contig47"  
89905. .95932  
/note="assembly\_name:Contig48"  
96033. .102531  
/note="assembly\_name:Contig49"  
102632. .111305  
/note="assembly\_name:Contig50"  
111406. .120462

```
repeat_region /rpt_family="ERV1"
repeat_region /rpt_family="ERV1"
repeat_region /rpt_family="ERV1"
misc_feature 10676..10831
misc_feature 10829..11215
misc_feature /note="similar to EST AW070186 (NID:g6025184) xa09a03.x1"
misc_feature 10829..11250
misc_feature /note="similar to EST AI806840 (NID:g59393406) wf15h08.x1"
repeat_region 11262..11805
repeat_region /rpt_family="ERV1"
misc_feature 12808..13247
misc_feature /note="similar to EST AI908793 (NID:g6499473)"
misc_feature 12809..12925
misc_feature /note="similar to EST AI908824 (NID:g6499504)"
repeat_region 12952..13235
repeat_region /rpt_family="ERV"
repeat_region 13237..13580
repeat_region /rpt_family="MaLR"
repeat_region 14241..14593
repeat_region /rpt_family="MaLR"
repeat_region 16242..16534
repeat_region /rpt_family="Alu"
misc_feature 17189..17450
misc_feature /note="similar to EST D62720 (NID:g966494)"
repeat_region 18011..18146
repeat_region /rpt_family="L1"
repeat_region 18705..18738
repeat_region /rpt_family="L1"
repeat_region 18739..20933
repeat_region /rpt_family="L1"
repeat_region 20934..21246
repeat_region /rpt_family="L1"
repeat_region 21247..21557
repeat_region /rpt_family="Alu"
repeat_region 21740..21881
repeat_region /rpt_family="Alu"
repeat_region 21983..22295
repeat_region /rpt_family="ERV1"
repeat_region 23119..23239
repeat_region /rpt_family="MIR"
repeat_region 23639..23934
repeat_region /rpt_family="L2"
repeat_region 24194..24335
repeat_region /rpt_family="MaLR"
repeat_region 24452..24562
repeat_region /rpt_family="L1"
repeat_region 24961..25528
repeat_region /rpt_family="ERVK"
repeat_region 25530..27490
repeat_region /rpt_family="ERVK"
repeat_region 27485..27807
repeat_region /rpt_family="ERV1"
repeat_region 27808..27914
repeat_region /rpt_family="ERV1"
repeat_region 27922..28018
repeat_region /rpt_family="ERV1"
repeat_region 27997..29043
repeat_region /rpt_family="ERV1"
repeat_region 31133..31165
repeat_region /rpt_family="ERVK"
repeat_region 31172..31597
repeat_region /rpt_family="ERVK"
repeat_region 31598..32237
repeat_region /rpt_family="ERV1"
repeat_region 34473..35285
repeat_region /rpt_family="L1"
repeat_region 35883..36498
repeat_region /rpt_family="ERV1"
repeat_region 38463..38929
repeat_region /rpt_family="MER1_type"
repeat_region 39354..39609
repeat_region /rpt_family="L1"

repeat_region 42594..42877
repeat_region /rpt_family="Alu"
repeat_region 43519..43817
repeat_region /rpt_family="Alu"
misc_feature 45998..46226
misc_feature /note="similar to EST AA909108 (NID:g3048513) om32d11.s1"
repeat_region 47950..48034
repeat_region /rpt_family="L1"
repeat_region 48869..48985
repeat_region /rpt_family="Alu"

Alignment Scores:
Pred. No.: 3.46e+03 Length: 117309
Score: 53.00 Matches: 11
Percent Similarity: 77.78% Conservative: 3
Best Local Similarity: 61.11% Mismatches: 4
Query Match: 41.09% Indels: 0
DB: 9 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x AC007274 (1-117309)
Oy 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluAlaAlaVal 21
Db 96844 CTGGAGGCTCTGCTGGACAGGACAGCGGGCACAGACAGGAGTAGCCTTG 96897

RESULT 19
AC087862/c
LOCUS AC087862 173772 bp DNA linear HTG 19-FEB-2001
DEFINITION Homo sapiens chromosome RPCI-11 clone RP11-724H13, WORKING DRAFT
SEQUENCE, 37 unordered pieces.
ACCESSION AC087862.3 GI:12963066
VERSION AC087862.3
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 173772)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 173772)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (02-FEB-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On Feb 18, 2001 this sequence version replaced gi:12745114.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0724H13
----- Summary Statistics -----
Sequencing vector: M13; 100%
Sequencing vector: plasmid; 0%
Chemistry: Dye-primer ET; 100% of reads
Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 150537 bases at least Q40
Consensus quality: 158418 bases at least Q30
Consensus quality: 162327 bases at least Q20
Insert size: 191000; agarose-fp
Insert size: 170172; sum-of-contigs
Quality coverage: 2.88 in Q20 bases; agarose-fp
Quality coverage: 3.33 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 37 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
```

```

/note="3 base direct repeat bordering transposon Tn5714.
Note difference - TCA in Streptomyces coelicolor plasmid
SCP2, TGA in Streptomyces coelicolor plasmid SCP1."
complement(7784..7794)
/note="Inverted and repeated imperfectly at
complement(11285..11295)"
complement(7784..11295)
/note="Similar to Tn5714 from SCP1"

Alignment Scores:
Pred. No.:      857      Length:      31317
Score:          53.00     Matches:      13
Percent Similarity: 65.38%   Conservative: 4
Best Local Similarity: 50.00% Mismatches: 9
Query Match:      41.09%   Indels:      0
DB:              1       Gaps:      0

US-09-847-539A-6_COPY_59_86 (1-28) x SCSCP2 (1-31317)
QY 3 AlaLeuGluAlaLeuAlaAspGInThrAspAlaLeuGlnSerGluGluAlaAlaVal 22
Db 23713 GCCGAGCGCGCGCGGACCGCGGAGCGCGGTCCAGTCGGAGCGCGCGCGCGGCC 23654
QY 23 LysAlaAspAsnAlaAla 28
Db 23653 GAAGCGGACCGCGCGGCC 23636

RESULT 18
AC007274
LOCUS      AC007274          117309 bp      DNA      linear      PRI 14-OCT-2000
DEFINITION Homo sapiens BAC clone RP11-105L10 from Y, complete sequence.
ACCESSION AC007274
VERSION   AC007274.3 GI:10801445
KEYWORDS  HTG.
SOURCE    human.
ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 117309)
AUTHORS   Sulston,J.E. and Waterston,R.
TITLE     Toward a complete human genome sequence
JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE   99063792
REFERENCE 2 (bases 1 to 117309)
AUTHORS   McPherson-Wagner,C., Maupin,R. and Krasucki,Peter.
TITLE     The sequence of Homo sapiens BAC clone RP11-105L10
JOURNAL   Unpublished
REFERENCE 3 (bases 1 to 117309)
AUTHORS   Waterston,R.H.
TITLE     Direct Submission
JOURNAL   Submitted (09-APR-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE 4 (bases 1 to 117309)
AUTHORS   Waterston,R.
TITLE     Direct Submission
JOURNAL   Submitted (21-DEC-1999) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE 5 (bases 1 to 117309)
AUTHORS   Waterston,R.
TITLE     Direct Submission
JOURNAL   Submitted (14-OCT-2000) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
            On Oct 14, 2000 this sequence version replaced gi:6624109.
            ----- Genome Center
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site: http://genome.wustl.edu/gsc
            Contact: saplens@wustl.wustl.edu
            ----- Summary Statistics
            -----
            Center project name: H_NH0105L10
            -----

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

#### MAPPING INFORMATION:

The position of this clone was established as part of a collaboration between the Human Chromosome Y Mapping Project (Tomoko Kawaguchi, Helen Skaletsky, Laura G. Brown, Steve Rozen, and David C. Page at the Whitehead Institute for Biomedical Research, Cambridge MA) and the Washington University Genome Sequencing Center, St. Louis MO.

#### SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E., Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBACe3.6

#### NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-305H21; the clone sequenced to the right is RP11-109F19, 200 bp overlap. Actual start of this clone is at base position 1 of RP11-105L10; actual end is at base position 24605 of RP11-109F19.

FEATURES	Source
repeat_region	1..117309
repeat_region	/organism="Homo sapiens"
repeat_region	/db_xref="taxon:9606"
repeat_region	/chromosome="Y"
repeat_region	/map="Y"
repeat_region	/clone="RP11-105L10"
repeat_region	/clone_lib="RPCI-11"
repeat_region	733..1035
repeat_region	/rpt_family="Alu"
repeat_region	1787..1857
repeat_region	/rpt_family="MIR"
repeat_region	1967..2286
repeat_region	/rpt_family="Alu"
repeat_region	3460..3864
repeat_region	/rpt_family="MAlR"
repeat_region	4654..4753
repeat_region	/rpt_family="MIR"
repeat_region	4917..5048
repeat_region	/rpt_family="MIR"
repeat_region	6100..6400
repeat_region	/rpt_family="Alu"
repeat_region	6920..7254
repeat_region	/rpt_family="L2"
repeat_region	7363..7578
repeat_region	/rpt_family="ERV"
repeat_region	7386..8065
repeat_region	/rpt_family="ERV1"
repeat_region	8066..8419
repeat_region	/rpt_family="ERV"
repeat_region	8762..8877
repeat_region	/rpt_family="ERV1"
repeat_region	8883..8945
repeat_region	/rpt_family="Alu"
repeat_region	8961..9178
repeat_region	/rpt_family="ERV1"
repeat_region	9185..10437





```
CDS
/db_xref="FLYBASE:FBgn0022943"
/join(4165..4212,4268..4485,4610..4808)
/gene="Cbp20"
/note="Cbp20 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0012357"
/db_xref="FLYBASE:FBgn0022943"
/protein_id="AAF55496.1"
/db_xref="GI:7300337"
/translation="MFASVELSYRDOHQFKSRSEQERSLRDSCITLVVGNLSFYTTEE
QIHFLSRGCDGVIVMGLDKYKTPCGFCFVEYVYRSEAEAMRVNCFRLDLRLR
VWDAGFVEGRQYGRGKTGGQVDEYRTDYDAGRGYKLLSKLIAPNDNR"
complement(join(4878..5331,5462..5596,5791..5935))
/gene="CG7217"
/product="CT22257"
/db_xref="FLYBASE:FBan0007217"
/db_xref="FLYBASE:FBgn0038570"
complement(<4878..>5935)
/gene="CG7217"
/map="90E3-90E4"
/db_xref="FLYBASE:FBan0007217"
/db_xref="FLYBASE:FBgn0038570"
complement(join(5002..5331,5462..5596,5791..5799))
/gene="CG7217"
/note="CG7217 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0007217"
/db_xref="FLYBASE:FBgn0038570"
/protein_id="AAF55497.1"
/db_xref="GI:7300338"
/translation="MKVGDLSFVSLFEDSPANKINTGDLVNGKKVIIFGVPGATFP
GCSKTHLPGVSSADELKSQGVDEIVCVSNDFPVMSANGKEHGAAGKVRLLADPAG
GFTKALDVTIDLPPLGCVRSKRYSLVYENGKVTLELVNVEPDGTGLSCSLANNIGKK"
complement(join(<6163..6507,6567..>6614))
/gene="CG7215"
/product="CT22253"
/db_xref="FLYBASE:FBan0007215"
/db_xref="FLYBASE:FBgn0038571"
complement(<6163..>6614)
/gene="CG7215"
/map="90E3-90E4"
/db_xref="FLYBASE:FBan0007215"
/db_xref="FLYBASE:FBgn0038571"
complement(join(6163..6507,6567..6614))
/gene="CG7215"
/note="CG7215 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0007215"
/db_xref="FLYBASE:FBgn0038571"
/protein_id="AAF55498.1"
/db_xref="GI:7300339"
/translation="MQITIKVLKGKDCITIEVAPTSTILEYKHQIEAELOISATNOKLL
LLGRPLANEQTIASYPNIKEGTKLNLVWIKPLRDSILRGFRKHSEPLAERTNPFM
ADPERKTNEQSLODLELADSVNRAAT"
/join(7165..7303,9093..9534,10965..>11147)
/gene="CG7901"
/product="CT23882"
/db_xref="FLYBASE:FBan0007901"
/db_xref="FLYBASE:FBgn0038572"
/join(7165..7303,8920..9534,10965..>11147)
/gene="CG7901"
/product="CT42545"
/db_xref="FLYBASE:FBan0007901"
/db_xref="FLYBASE:FBgn0038572"
<7165..>11147
/gene="CG7901"
/map="90E3-90E5"
/db_xref="FLYBASE:FBan0007901"
/db_xref="FLYBASE:FBgn0038572"
/join(9142..9534,10965..11147)
/gene="CG7901"
/note="CG7901 gene product [alt 1]"
/codon_start=1
```

```
CDS
/db_xref="FLYBASE:FBan0007901"
/db_xref="FLYBASE:FBgn0038572"
/protein_id="AAF55499.1"
/db_xref="GI:7300340"
/translation="MDNEALDPTIKSSTAATPAASETTTTTAASSVVETTITAAAT
TASAEKNETTATNNSNTSGSISSSSNNIVIPASATNGIKESNLSLTITTTAAAY
AAATTEGVAPAITSTIVTGGTPPLSSLANLKLDNTPPYDAPPPPIISKVLNITGTP
IVRKERQTSARYNASKNCELTALIPLENE"
/join(9142..9534,10965..11147)
/gene="CG7901"
/note="CG7901 gene product [alt 2]"
/codon_start=1
/db_xref="FLYBASE:FBan0007901"
/db_xref="FLYBASE:FBgn0038572"
/protein_id="AAF55500.1"
/db_xref="GI:7300341"
/translation="MDNEALDPTIKSSTAATPAASETTTTTAASSVVETTITAAAT
TASAEKNETTATNNSNTSGSISSSSNNIVIPASATNGIKESNLSLTITTTAAAY
AAATTEGVAPAITSTIVTGGTPPLSSLANLKLDNTPPYDAPPPPIISKVLNITGTP
IVRKERQTSARYNASKNCELTALIPLENE"
/join(12280..12385,12562..13626,13976..14480,14769..14965,
15039..15373,15432..16135,16241..16677)
/gene="CG7913"
/product="CT23900"

*
mRNA
US-09-847-539a-6_COPY_59_86 (1-28) x AE003721 (1-224896)

Qy 5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaValValLysAla 24
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 13577 GAGGAGGAAGCTGACGAGGAGGACGCCGCCGAGGATGCGCGCGTGC CGCC 13518

Qy 25 AspAsnAlaAla 28
||| |||||

Db 13517 GATGCTCGGCGG 13506

RESULT 16
SCO414671 7686 bp DNA circular BCT 19-DEC-2001
LOCUS
DEFINITION Streptomyces coelicolor plasmid 2 (SCP2*).
ACCESSION AJ414671
VERSION AJ414671.1 GI:17974213
KEYWORDS ORF1; ORF2; ORF3; ORF4; ORF5; ORF6.
SOURCE Streptomyces coelicolor.
ORGANISM Streptomyces coelicolor
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (sites)
AUTHORS Wilkenson,C.J., Hughes-Thomas,Z.A., Martin (nee Rowe),C.J.,
Bohm,I., Mironenko,T., Deacon,M., Wheatcroft,M., Wirtz,G.,
Staunton,J. and Leadlay,P.F.
Increasing the efficiency of heterologous promoters in
actinomycetes
J. Mol. Microbiol. Biotechnol. In press
REFERENCE 2 (bases 1 to 7686)
AUTHORS Wilkenson,C.J.
TITLE Direct Submission
JOURNAL Submitted (28-SEP-2001) Wilkenson C.J., Department of Biochemistry,
University of Cambridge, 80 Tennis Court Rd., Cambridge, CB2 1GA,
UNITED KINGDOM
FEATURES
source
1..7686
/organism="Streptomyces coelicolor"
/plasmid="2 SCP2*"
/db_xref="taxon:1902"
1261..1713

CDS
```

\* 152667 153590: contig of 924 bp in length  
 \* 153591 153670: gap of unknown length  
 \* 153671 154319: contig of 649 bp in length  
 \* 154320 154399: gap of unknown length  
 \* 154400 154743: contig of 344 bp in length  
 \* 154744 154823: gap of unknown length  
 \* 154824 155222: contig of 599 bp in length  
 \* 155223 155502: gap of unknown length  
 \* 155503 155840: contig of 338 bp in length.

## FEATURES

source

1. .155840  
 /organism="Drosophila melanogaster"  
 /strain="y: cn bw sp"  
 /db\_xref="taxon:7227"  
 /chromosome="3"  
 /map="90E1-90F10"  
 /clone\_lib="BAC48113 (D522) RPCI-98 48.I.13"  
 /clone\_lib="RPCI-98 (Roswell Park Cancer Institute  
 Drosophila melanogaster BAC library, partial EcoRI in  
 pRAC3.6"

BASE COUNT 41735 a 34360 c 34802 g 41426 t 3517 others  
 ORIGIN

## Alignment Scores:

Pred. No.: 3,23e+03 Length: 155840  
 Score: 54.00 Matches: 12  
 Percent Similarity: 66.67% Conservatives: 4  
 Best Local Similarity: 50.00% Mismatches: 8  
 Query Match: 41.86% Indels: 0  
 DB: 2 Gaps: 0

US-09-847-539a-6\_COPY\_59\_86 (1-28) x AC006495 (1-155840)

QY 5 GluAlaLeuAlaSpGlnThrAspAlaLeuGlnSerGluAlaAlaValValysAla 24  
 ||| |||||: ||||| :|||: ||||| |||

Db 13996 GAGGAGGAAGCTGACGAGGAGGAGCGCGCGAGGAGGATCGCGCGCGCG 13937  
 ||| |||||: ||||| :|||: ||||| |||

QY 25 AspAsnAlaAla 28  
 ||| |||||

Db 13936 GATGCTGCGCGG 13925

## RESULT 15

AE003721/c

LOCUS 224896 bp DNA linear INV 05-OCT-2000  
 Drosophila melanogaster genomic scaffold 142000013386035 section 46  
 of 105, complete sequence.

ACCESSION AE003721 AE002708

VERSION AE003721.1 GI:7300335

KEYWORDS HTG.

SOURCE fruit fly.

ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 224896)

ADAMSON, M. D., CELINKER, S. E., HOLT, R. A., EVANS, C. A., GOCAYNE, J. D.,  
 GEORGE, R. A., LEWIS, S. E., RICHARDS, S., ASHBURNER, M., HENDERSON, S. N.,  
 SUTTON, G. G., WORTMAN, J. R., YANDELL, M. D., ZHANG, Q., CHEN, L. X.,  
 BRANDON, R. C., ROGERS, Y. H., BLAZEJ, R. G., CHAMPE, M., PFELFER, B. D.,  
 WAN, K. H., DOYLE, C., BAXTER, E. G., HELT, G., NELSON, C. R., GABOR  
 MIKLOS, G. L., ABRIL, J. F., AGHAYANI, A., AN, H. J., BASU, A.,  
 ANDREWS, PFANNKUCH, C., BALDWIN, D., BALLEW, R. M., BEESON, K. Y.,  
 BAXENDALE, D., BAYRAKCI, L., BEASLEY, E. M., BEESON, K. Y.,  
 BENOS, P. V., BERMAN, B. P., BHANDARI, D., BOISHAKOV, S., BORKOVA, D.,  
 BOTCHAN, M. R., BOUCK, J., BROKSTEIN, P., BROTTIER, P., BURTON, K. C.,  
 BUSAM, D. A., BUTLER, H., CADIEU, E., CENTER, A., CHANDRA, I.,  
 CHERRY, J. M., CAWLEY, S., DAHLE, C., DAVENPORT, L. B., DAVIES, P., DE  
 PABLOS, B., DELCHER, A., DENG, Z., MAYS, A. D., DEW, I., DIETZ, S. M.,  
 DODSON, K., DOUP, L. E., DOWNES, M., DUGAN-ROCHA, S., DUNKOV, B. C.,  
 DUNN, P., DURBIN, K. J., EVANGELISTA, C. C., FERRAZ, C., FERRIERA, S.,  
 FLEISCHMANN, W., FOSLER, C., GABRIELIAN, A. E., GARG, N. S.,  
 GELBART, W. M., GLASSER, K., GLODEK, A., GONG, F., GORRELL, J. H., GU, Z.,

Guan, P., Harris, M., Harris, N. L., Harvey, D., Helman, T. J.,  
 Hernandez, J. R., Houck, J., Hostin, D., Houston, K. A., Howland, T. J.,  
 Wei, M. H., Ibegwam, C., Jalali, M., Kalush, F., Karpen, G. H., Ke, Z.,  
 Kennison, J. A., Ketchum, K. A., Kimmel, B. E., Kodira, C. D., Kraft, C.,  
 Kravitz, S., Kulp, D., Lai, Z., Lasko, P., Lei, Y., Levitsky, A. A.,  
 Li, J., Li, Z., Liang, Y., Lin, X., Liu, X., Mattel, B., McIntosh, R. C.,  
 McLeod, M. P., McPherson, D., Merkulov, G., Milahina, N. V., Mobarry, C.,  
 Morris, J., Moshrefi, A., Mount, S. M., Moy, M., Murphy, B., Murphy, L.,  
 Muzny, D. M., Nelson, D. L., Nelson, D. R., Nelson, K. A., Nixon, K.,  
 Nusskern, D. R., Pacleb, J. M., Palazzolo, M., Pittman, G. S., Pan, S.,  
 Pollard, J., Puri, V., Reese, M. G., Reinert, K., Remington, K.,  
 Saunders, R. D., Scheeler, F., Shen, H., Shue, B. C., Siden-Kiamos, I.,  
 Simpson, M., Skupski, M. P., Smith, T., Spier, E., Spradling, A. C.,  
 Stapleton, M., Strong, R., Sun, E., Svirskas, R., Tector, C., Turner, R.,  
 Venter, E., Wang, A. H., Wang, X., Wang, Z. Y., Wasserman, D. A.,  
 Weinstein, G. M., Weissbach, J., Williams, S. M., Woodage, T.,  
 Worley, K. C., Wu, D., Yang, S., Yao, Q. A., Ye, J., Yeh, R. F.,  
 Zaveri, J. S., Zhan, M., Zhang, G., Zhao, Q., Zheng, L., Zheng, X. H.,  
 Zhong, F. N., Zhong, W., Zhou, X., Zhu, S., Zhu, X., Smith, H. O.,  
 Gibbs, R. A., Myers, E. W., Rubin, G. M. and Venter, J. C.  
 The genome sequence of *Drosophila melanogaster*  
 Science 287 (5461), 2185-2195 (2000)  
 20196006  
 2 (bases 1 to 224896)  
 Adams, M. D., Celinker, S. E., Gibbs, R. A., Rubin, G. M. and Venter, J. C.  
 Direct Submission  
 Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,  
 Rockville, MD, USA  
 Location/Qualifiers  
 1. .224896  
 /organism="Drosophila melanogaster"  
 /db\_xref="taxon:7227"  
 /chromosome="3R"  
 complement(join(1202. .2465,2540. .3150,3540. .3806))  
 /gene="CG7218"  
 /product="CT22261"  
 /db\_xref="FLYBASE:FBan0007218"  
 /db\_xref="FLYBASE:FBgn0038569"  
 complement(<1202. .>3806)  
 /gene="CG7218"  
 /map="90E2-90E4"  
 /db\_xref="FLYBASE:FBan0007218"  
 /db\_xref="FLYBASE:FBgn0038569"  
 complement(join(1245. .2465,2540. .3150,3540. .3738))  
 /gene="CG7218"  
 /note="CG7218 gene product"  
 /codon\_start=1  
 /db\_xref="FLYBASE:FBan0007218"  
 /db\_xref="FLYBASE:FBgn0038569"  
 /protein\_id="AAF55495.2"  
 /db\_xref="GI:10726596"  
 /translation="MNATNSAGGRKQLRFRGDTGSRVEELHQOQEKQKAPLQ  
 DDVAATPTATPAAGSAQRLQSGTACTNTFTDFPKVEMTRCYMLEHDEERYSARRQ  
 KIYSFMRIPIRLERFMVYIMQCADSFYIHTFLPFRVFMVAALVSRVARIFLRLS  
 SGRLSPAECIDLLKGVITWTVTLIMLLADVNRVHIIKSQSIKLYTFYKNLEVG  
 RLNSAFGQDTIDALFWTATEPKNSREHFGVLTHTVLTLYVFLSGLMFOATCLNV  
 AVNSAGQGLTIMSNFVELGSKFVKDKNNLFQLTCDVRERPHLSVLTFTVYIQ  
 TKRQKHSITQFCVMLPCDFAVLTETLIDWVKHAPITFNPESLPSIYREYTSUAYD  
 MTQTKDFASFDLSVLARRMGFIPTPLAVVLIKATYAVSFENAAWLLFLAYLFA  
 MGLRCLTICALGKACAKKEHQTERNSPTSSMTNPVPIGAAAPVSAATGQNHNN  
 NNNNNNSISIGSKPAQVTTLLTPPSAGHLDSYKNSFTSIATSTPTPKAVSEDLV  
 TNSLELGATLPSNSDVLDDVCLNEQVNTNTSSAEOVQEDQLVRSOPDLMLNN  
 SGQGVTSAAKAKATQRLPRTKRSSEPCMSWKEGGAAGIAGNGQTTL".  
 join(4128. .4212,4268. .4485,4610. .4873)  
 /gene="Cbp20"  
 /product="CT23846"  
 /db\_xref="FLYBASE:FBan0012357"  
 /db\_xref="FLYBASE:FBgn0022943"  
 <4128. .>4873  
 /gene="Cbp20"  
 /note="CG12357"  
 /map="90E3-90E4"  
 /db\_xref="FLYBASE:FBan0012357"

mRNA

gene



```
/db_xref="taxon:7227"
/chromosome="3R"
```

/chromosome="3R"

map="90E-90E"

```
/clone="BACR27G04 (D985)"
```

/clone\_lib="RPCI-98 (Roswell Park Cancer Institute  
Drosophila melanogaster BAC library, partial EcoRI in  
pBAC3.6)"

BASE COUNT	43139 a	33567 c	32376 g	42528 t
------------	---------	---------	---------	---------

## ORIGIN

**Alignment Scores:**

Pred. No.:	3.14e+03	Length:	151610
Score:	54.00	Matches:	12
Percent Similarity:	66.67%	Conservative:	4
Best Local Similarity:	50.00%	Mismatches:	8
Query Match:	41.86%	Indels:	0
DB:	3	Gaps:	0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x AC009462 (1-151610)

Qy 5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluAlaAlaValLysAla 24

[illegible]

Db 73153 GAGGAGGAAGCTGACGAGGAGGACGCCGCCGAGGCAGAGGATGCCGCGCTGCCGCC 73094

Qy 25 AspAsnAlaAla 28

—

Db 73093 GATGCTGCGGCG 73082

## RESULT 14

AC006495/C

**LOCUS**

DEFINITION Drosophila melanogaster chromosome 3 clone BACR48113 (D522) 48.1.13 map 90E1-90F10 strain Y; cn bw sp, \*\*\* SEQUENCING IN PROGRESS \*\*\* 38 unordered pieces.

ACCESSION AC006495

VERSION AC006495.17 GI:6838793

**KEYWORDS** HTG: HTGS PHASE1.

SOURCE	fruit fly.
1. <i>Phaenocarpa</i>	
2. <i>Phaenocarpa</i>	
3. <i>Phaenocarpa</i>	
4. <i>Phaenocarpa</i>	
5. <i>Phaenocarpa</i>	
6. <i>Phaenocarpa</i>	
7. <i>Phaenocarpa</i>	
8. <i>Phaenocarpa</i>	
9. <i>Phaenocarpa</i>	
10. <i>Phaenocarpa</i>	
11. <i>Phaenocarpa</i>	
12. <i>Phaenocarpa</i>	
13. <i>Phaenocarpa</i>	
14. <i>Phaenocarpa</i>	
15. <i>Phaenocarpa</i>	
16. <i>Phaenocarpa</i>	
17. <i>Phaenocarpa</i>	
18. <i>Phaenocarpa</i>	
19. <i>Phaenocarpa</i>	
20. <i>Phaenocarpa</i>	
21. <i>Phaenocarpa</i>	
22. <i>Phaenocarpa</i>	
23. <i>Phaenocarpa</i>	
24. <i>Phaenocarpa</i>	
25. <i>Phaenocarpa</i>	
26. <i>Phaenocarpa</i>	
27. <i>Phaenocarpa</i>	
28. <i>Phaenocarpa</i>	
29. <i>Phaenocarpa</i>	
30. <i>Phaenocarpa</i>	
31. <i>Phaenocarpa</i>	
32. <i>Phaenocarpa</i>	
33. <i>Phaenocarpa</i>	
34. <i>Phaenocarpa</i>	
35. <i>Phaenocarpa</i>	
36. <i>Phaenocarpa</i>	
37. <i>Phaenocarpa</i>	
38. <i>Phaenocarpa</i>	
39. <i>Phaenocarpa</i>	
40. <i>Phaenocarpa</i>	
41. <i>Phaenocarpa</i>	
42. <i>Phaenocarpa</i>	
43. <i>Phaenocarpa</i>	
44. <i>Phaenocarpa</i>	
45. <i>Phaenocarpa</i>	
46. <i>Phaenocarpa</i>	
47. <i>Phaenocarpa</i>	
48. <i>Phaenocarpa</i>	
49. <i>Phaenocarpa</i>	
50. <i>Phaenocarpa</i>	
51. <i>Phaenocarpa</i>	
52. <i>Phaenocarpa</i>	
53. <i>Phaenocarpa</i>	
54. <i>Phaenocarpa</i>	
55. <i>Phaenocarpa</i>	
56. <i>Phaenocarpa</i>	
57. <i>Phaenocarpa</i>	
58. <i>Phaenocarpa</i>	
59. <i>Phaenocarpa</i>	
60. <i>Phaenocarpa</i>	
61. <i>Phaenocarpa</i>	
62. <i>Phaenocarpa</i>	
63. <i>Phaenocarpa</i>	
64. <i>Phaenocarpa</i>	
65. <i>Phaenocarpa</i>	
66. <i>Phaenocarpa</i>	
67. <i>Phaenocarpa</i>	
68. <i>Phaenocarpa</i>	
69. <i>Phaenocarpa</i>	
70. <i>Phaenocarpa</i>	
71. <i>Phaenocarpa</i>	
72. <i>Phaenocarpa</i>	
73. <i>Phaenocarpa</i>	
74. <i>Phaenocarpa</i>	
75. <i>Phaenocarpa</i>	
76. <i>Phaenocarpa</i>	
77. <i>Phaenocarpa</i>	
78. <i>Phaenocarpa</i>	
79. <i>Phaenocarpa</i>	
80. <i>Phaenocarpa</i>	
81. <i>Phaenocarpa</i>	
82. <i>Phaenocarpa</i>	
83. <i>Phaenocarpa</i>	
84. <i>Phaenocarpa</i>	
85. <i>Phaenocarpa</i>	
86. <i>Phaenocarpa</i>	
87. <i>Phaenocarpa</i>	
88. <i>Phaenocarpa</i>	
89. <i>Phaenocarpa</i>	
90. <i>Phaenocarpa</i>	
91. <i>Phaenocarpa</i>	
92. <i>Phaenocarpa</i>	
93. <i>Phaenocarpa</i>	
94. <i>Phaenocarpa</i>	
95. <i>Phaenocarpa</i>	
96. <i>Phaenocarpa</i>	
97. <i>Phaenocarpa</i>	
98. <i>Phaenocarpa</i>	
99. <i>Phaenocarpa</i>	
100. <i>Phaenocarpa</i>	

ORGANISM *Drosophila melanogaster*

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

## REFERENCE

## AUTHORS

**TITLE**

JOURNAL

REF ID: A66862

## AUTHORS

**DATE**

## ATTI

JOURNAL

COMMENT

Laboratory, MS 64-121, Berkeley, CA 94720, USA  
On Jan 31, 2003 this sequence version replaced gi:5687780.  
For further information about this sequence, including its location  
and relationship to other sequences, please visit our sequence  
archive web site (<http://www.fruitfly.org/sequence/>) or send email  
to [bdgs@fruitfly.berkeley.edu](mailto:bdgs@fruitfly.berkeley.edu). All contigs in this submission meet  
the following cutoffs: length > 200 bases.  
\* NOTE: This is a 'working draft' sequence. It currently

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 38 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

667:	contig of 667 bp in length
747:	gap of unknown length
1288:	contig of 541 bp in length
1368:	gap of unknown length
7705:	contig of 6337 bp in length
7785:	gap of unknown length
13266:	contig of 5481 bp in length
13346:	gap of unknown length
134516:	contig of 121170 bp in length
134596:	gap of unknown length
135458:	contig of 862 bp in length
135358:	gap of unknown length
135764:	contig of 226 bp in length
135844:	gap of unknown length
136080:	contig of 236 bp in length
136160:	gap of unknown length
136683:	contig of 523 bp in length
136763:	gap of unknown length
137566:	contig of 803 bp in length
137646:	gap of unknown length
138152:	contig of 506 bp in length
138232:	gap of unknown length
138485:	contig of 253 bp in length
138565:	gap of unknown length
139286:	contig of 721 bp in length
139366:	gap of unknown length
140013:	contig of 647 bp in length
140093:	gap of unknown length
140815:	contig of 722 bp in length
140895:	gap of unknown length
141513:	contig of 618 bp in length
141593:	gap of unknown length
142372:	contig of 779 bp in length
142452:	gap of unknown length
143112:	contig of 660 bp in length
143192:	gap of unknown length
144042:	contig of 850 bp in length
144122:	gap of unknown length
144683:	contig of 561 bp in length
144763:	gap of unknown length
145169:	contig of 406 bp in length
145249:	gap of unknown length
146124:	contig of 875 bp in length
146204:	gap of unknown length
146979:	contig of 775 bp in length
147059:	gap of unknown length
147581:	contig of 522 bp in length
147661:	gap of unknown length
148202:	contig of 541 bp in length
148282:	gap of unknown length
148505:	contig of 223 bp in length
148585:	gap of unknown length
149025:	contig of 440 bp in length
149105:	gap of unknown length
149495:	contig of 390 bp in length
149575:	gap of unknown length
150161:	contig of 586 bp in length
150241:	gap of unknown length
150736:	contig of 495 bp in length
150816:	gap of unknown length
151159:	contig of 343 bp in length
151239:	gap of unknown length
152123:	contig of 884 bp in length
152203:	gap of unknown length
152586:	contig of 383 bp in length
152666:	gap of unknown length

```
/gene="GRL5"  
/note="Ca2+ binding protein; translated as a proprotein;  
Grl5p"  
/codon_start=1  
/transl_table=6  
/product="granule lattice protein 5 precursor"  
/protein_id="AAC27989.1"  
/db_xref="GI:3366655"  
/translation="MRVIALLVIALVLCOSAVTSSKQAKLMEKINSKLEKSHLGR  
ALKNWTATKGYQDYLDYDAFAALKNOLLNLDNENSLFETOTASHDSVAQFNAD  
ISNGYQIDAQGLDNLSDNTYQONLDQAQALQDNDTALNAEAEALNAEALYQ  
VATRYSDADQVIGLAEKQESHYDNADLGSFSFQIKNTFSFAQVTEATKGM  
NAKHQLEFVPVQAMQVKNNTSQSSITAIKALQDQAYFKFTFSDLNEYVFTQN  
VOSTDGLNQITDILQNOVLPGEAQISLQAOIQOVEDALALAAQQLDAQNALDAE  
NAQWENVARHQALVDRINSEYDLVTQADRVNQAQVNGSD"  
mat_peptide  
195..1134
```

```
/gene="GRL5"
```

```
/note="Grl5p"
```

```
/product="granule lattice protein 5"
```

```
BASE COUNT 417 a 314 c 202 g 355 t
```

ORIGIN

#### Alignment Scores:

Pred. No.:	20.5	Length:	1288
Score:	54.00	Matches:	12
Percent Similarity:	66.67%	Conservative:	4
Best Local Similarity:	50.00%	Mismatches:	8
Query Match:	41.86%	Indels:	0
DB:	3	Gaps:	0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x AF031321 (1-1288)

QY 2 AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaVal 21

Db 420 GATGCTCAACAGCTCTTAAGACACACTGATGCTCTCAATGCTGCCGAAGACTCTT 479

QY 22 VallysAlaAsp 25

Db 480 GCCAATGCTGA 491

#### RESULT 12

AC014470/c

LOCUS

DEFINITION Drosophila melanogaster, \*\*\* SEQUENCING IN PROGRESS \*\*\*, in ordered

pieces.

AC014470

VERSION AC014470.1 GI:6436865

KEYWORDS HTG; HTGS\_PHASE2.

SOURCE fruit fly.

ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 32270)

AUTHORS Adams, M. and Venter, J.C.

TITLE Direct Submission

JOURNAL Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,

Rockville, MD, USA

COMMENT This sequence was identified as CDM:10212984 by the submitter.

\* NOTE: This is a 'working draft' sequence.

\* This sequence will be replaced

\* by the finished sequence as soon as it is available and

\* the accession number will be preserved.

#### FEATURES

source

1..32270

/organism="Drosophila melanogaster"

/db\_xref="taxon:7227"

BASE COUNT 9248 a 7110 c 6761 g 9151 t

ORIGIN

#### Alignment Scores:

Pred. No.:

613

Length:

32270

Score: 54.00 Matches: 12  
Percent Similarity: 66.67% Conservative: 4  
Best Local Similarity: 50.00% Mismatches: 8  
Query Match: 41.86% Indels: 0  
DB: 2 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x AC014470 (1-32270)

QY 5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValLysAla 24

Db 26993 GAGGAGGAAGCTGACGAGGAGGAGCGCCGCGAGGAGAGGCTGCCGCC 26934

QY 25 AspAsnAlaAla 28

Db 26933 GATGCTCGCGG 26922

#### RESULT 13

AC009462/c

LOCUS

DEFINITION

AC009462

VERSION AC009462.6 GI:15451514

KEYWORDS HTG.

SOURCE fruit fly.

ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 151610)

AUTHORS

Celniker, S.E., Adams, M.D., Kronmiller, B., Tyler, D., Wan, K.H.,

Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Brandon, R.C.,

Rogers, Y., An, H., Baidwin, D., Banzon, J., Beeson, K.Y., Busam, D.A.,

Carlson, J.W., Center, A., Champe, M., Davenport, L.B., Dietz, S.M.,

Dodson, K., Dorsett, V., Dou, L.E., Doyle, C., Dresnek, D., Farfan, D.,

Ferrier, S., Frise, E., Galle, R.F., Garg, N.S., George, R.A.,

Gonzalez, M., Houck, J., Hoskins, R.A., Hostin, D., Howland, T.J.,

Ibegwam, C., Jalali, M., Kruse, D., Li, P., Mattel, B., Moshrefi, A.,

McIntosh, T.C., Moy, M., Murphy, B., Nelson, C., Nelson, K.A., Nunoo, J.,

Paclet, J., Paragas, V., Park, S., Patel, S., Pfeiffer, B.,

Phuanavanong, S., Pittman, G.S., Puri, V., Richards, S., Scheeler, F.,

Shapleton, M., Strong, R., Svirskas, R., Tector, C., Williams, S.M.,

Zaveri, J.S., Smith, H.O., Rubin, G.M. and Venter, J.C.

Sequencing of Drosophila chromosome 3R, region 90E-90E

Unpublished

2 (bases 1 to 151610)

Celniker, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blazej, R.G.,

Butenhoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L.,

Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,

Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L.,

Kim, E., Lee, B., Lewis, S., Li, P., Lomotan, M.A., Mazda, P.,

Moshrefi, A.R., Moshrefi, M., Nixon, K., Paclet, J.M., Park, S.,

Pfeiffer, B., Poon, L., Sequeira, A., Sethi, H., Snir, E.,

Svirskas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and

Rubin, G.M.

Direct Submission

Submitted (23-AUG-1999) Drosophila Genome Center, Lawrence Berkeley

Laboratory, MS 64-121, Berkeley, CA 94720, USA

On Sep 6, 2001 this sequence version replaced gi:13384332.

Sequence submitted by:

Berkeley Drosophila Genome Project

Lawrence Berkeley National Laboratory, MS 64-121

Berkeley, CA 94720

This sequence was assembled using end sequences from a whole genome

shotgun and from subclones of this BAC and its neighboring clones.

For further information about this sequence, including its location

and relationship to other sequences, please visit our sequence

archive web site (<http://www.fruitfly.org/sequence/>) or send email

to [bdg@fruitfly.berkeley.edu](mailto:bdg@fruitfly.berkeley.edu).

Location/Qualifiers

1..151610

/organism="Drosophila melanogaster"

/strain="y; cn bw sp"



```
/codon_start=1
/transl_table=11
/product="CATION EFFLUX SYSTEM PROTEIN"
/protein_id="A006822.1"
/db_xref="GI:4155849"
/translation="MIEKIIDLVSNNKLLTTLVTLIFLASLWAKSVRLDALPDLSLP
AQVVDITVPNQSPKIVQEQVTPVLSYFMSIANIDTVRGISSEGLIYIIFKDGYN
LYWADRVLEQLNRVNLKPKDAKVEIGSOSTSIGWAYQALSDSKNLSLKVLPQFY
YRVALGQVGSVSVGGVGVKDYEVTVLQNDLSIRYNLSLEOVANAIPKSNNDTGGV
ILENGEKIIRSHGYSIQSLNDLEIYVKKEGATPLAKIDIASVRLAPKPRGAANLNG
DKEVGGINVRHADTYKVKAIKEKATLQASNPDKITSVIDRSELKEGIDNLI
HTLIEESVIVLIIAIFLHRSALVVIITPLSVSCISFLMRFYNEIASIMLGGTIA
IATGAMVDAIVMVENAHKHLQIDTRDNTQVNAIQGVKVGGAIFFAIMLIVVSF
LPIFALTGOBEKFLAPLAYTKTFAMVLGALLSITIVPLMWLIIKGRILBESKNPNA
FPMKIVGSVLKVKRYAPFLIASVLGGLGVAYKKLWEEFIPQINEGVIMYTPVTI
NGVIGIDTALEYLAKSNAAKOLDVFKVQVCKGRANTSDAAGLMEITYELKPNQE
WKEKLSYKEVRDKLETKLQGLTNSNYITPIRGTDMLJTGIRTPGKIUGNDTKL
QKAILMEQOKLTKESLSVFAERSNNYYITLDNDENLARYGINKNAVLDITKAL
GGATLTMTIKGVESYISLREDTERTNTEKLNLYIKTAAYMPLRELARHYIDRSP
AVLSEKGLNVNFIYIVPQNGISSDITVROLAOKLEKIQLPSCGYVYFEFSGSOYLEEA
FKTLQIVPVSVFIIFLIVFALKNLNTSLCFFTLPPFAFLGLGFIEMNLGFENSVAA
```

## Alignment Scores:

Pred. No.:	75	Length:	12526
Score:	57.00	Matches:	12
Percent Similarity:	69.57%	Conservative:	4
Best Local Similarity:	52.17%	Mismatches:	7
Query Match:	44.19%	Indels:	0
DB:	1	Gaps:	0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x AE001548 (1-12526)

```
Qy 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVallys 23
Db 5073 ATACAGCCCTACAGGACGAATTGACGCTTTAGATTCTCAAGAAAAGTGCTTAGCAAA 5132
Qy 24 AlaAspAsn 26
Db 5133 TGGGATAAC 5141
|||||
```

```
RESULT 10
STYPPHMC2/c
LOCUS Salmonella enterica serovar Typhi (Salmonella typhi) strain Crl8
DEFINITION plasmid pHCM2.
ACCESSION AL513384
VERSION AU513384.1 GI:16505981
```

```
SOURCE Salmonella enterica subsp. enterica serovar Typhi.
ORGANISM Salmonella enterica subsp. enterica serovar Typhi
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Salmonella.
```

```
REFERENCE 1 (bases 1 to 106516)
AUTHORS Parkhill,J., Dougan,G., James,K.D., Thomson,N.R., Pickard,D.,
Wain,J., Churcher,C., Mungall,K.L., Bentley,S.D., Holden,M.T.G.,
Sebahia,M., Baker,S., Basham,D., Brooks,K., Chillingworth,T.,
Conerton,P., Cronin,A., Davis,P., Davies,R.M., Dowd,L., White,N.,
Farrar,J., Feltwell,T., Hamlin,N., Haque,A., Hien,T.T., Holroyd,S.,
Jagels,K., Krogh,A., Larsen,T.S., Leather,S., Moule,S., O'Gaora,P.,
Parry,C., Quail,M., Rutherford,K., Simmonds,M., Skelton,J.,
Stevens,K., Whitehead,S. and Barrell,B.G.
Complete genome sequence of a multiple drug resistant Salmonella
enterica serovar Typhi Crl8
```

```
JOURNAL Nature 413 (6858), 848-852 (2001)
PUBMED 11677608
REFERENCE 2 (bases 1 to 106516)
AUTHORS Parkhill,J.
TITLE Direct Submission
```

```
JOURNAL Submitted (25-OCT-2001) Submitted on behalf of the Salmonella
sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk
Notes:
COMMENT Details of S. typhi sequencing at the Sanger Centre are available
```

on the World Wide Web.  
(URL, [http://www.sanger.ac.uk/Projects/S\\_typhi/](http://www.sanger.ac.uk/Projects/S_typhi/)).

## FEATURES

source	1..106516
organism	"Salmonella enterica subsp. enterica serovar Typhi"
plasmid	"pHCM2"
strain	"CT18"
db_xref	"taxon:90370"
complement	(1..29)
gene	/gene="HCM2.0001c"
CDS	complement(1..29) /partial /gene="HCM2.0001c" highly similar to TR:CA855227 (EMBL:AL117211), YPWT1.45C, Yersinia pestis CO-92 hypothetical protein from plasmid pMT1 (71 aa), fasta scores: E(): 5.9e-25, 93.3% identity in 60 aa overlap and TR:O68760 (EMBL:AF074611), Y1119, Yersinia pestis KIM5 hypothetical protein from plasmid pMT1 (71 aa), fasta scores: E(): 5.9e-25, 93.3% identity in 60 aa overlap. Spans the end of the sequence"
misc_feature	1..232 /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="CAD09869.1" /db_xref="GI:16505982" /translation="MAVTLAGLE"
gene	/note="95.7% identical to Yersinia pestis CO-92 plasmid pMT1 (EMBL:AL117211) bases 48304..48535" complement(29..334) /gene="HCM2.0002c" complement(29..334) /gene="HCM2.0002c"
CDS	/note="HCM2.0002c, possible periplasmic protein, len: 101 aa; similar to TR:CA855228 (EMBL:AL117211), YPWT1.46C, Yersinia pestis hypothetical protein from plasmid pMT1 (111 aa), fasta scores: E(): 2.8e-26, 95.5% identity in 67 aa overlap. The N-terminal 37 aa are highly similar to YPMT1.46AC (EMBL:AL117211) Yersinia pestis hypothetical protein from plasmid pMT1 (59 aa) (86.8% identity in 38 aa overlap) which lies immediately upstream of YPMT1.46C in pMT1. Contains a possible N-terminal signal sequence"
misc_feature	/codon_start=1 /transl_table=11 /product="putative periplasmic protein" /protein_id="CAD09869.1" /db_xref="GI:16505983" /translation="MKTVDLTLLIVSIAFVLDICIFTGATRKALAPVNGAMVVDSGEQ FYVVRDRLSSPYLTKRNGKLSGVGEDKFKVYKSGDVYGVHAKNASYLPDDCKEVG"
RBS	complement(36..43) /gene="HCM2.0002c" /note="possible RBS"
misc_feature	224..6722 /note="96.5% identical to Yersinia pestis CO-92 plasmid pMT1 (EMBL:AL117211) bases 48733..55208"
gene	complement(375..650) /gene="HCM2.0003c"
CDS	complement(375..650) /gene="HCM2.0003c" /note="HCM2.0003c, highly similar to TR:CA862371 (EMBL:AL117211), YPWT1.46BC, Yersinia pestis CO-92 hypothetical protein from plasmid pMT1 (91 aa), fasta scores: E(): 0, 94.5% identity in 91 aa overlap and TR:Q92GX9 (EMBL:AF074611), Y1117, Yersinia pestis KIM5 hypothetical protein from plasmid pMT1 (91 aa), fasta scores: E(): 0, 94.5% identity in 91 aa overlap"
misc_feature	/codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="CAD09870.1"



```

/translation="MNPEKATHCKEIRKILKEALHNDKDLNLYLESSGSHAKLTSGAN
HLIPSSDRKSVKNFEKELTEFKRLRENTHITETCE"
complement(2949..4340)
/gene="HP1325"
CDS
complement(2949..4340)
/gene="HP1325"
/notes="similar to SP:P05042 GB:X00522 GB:X04065 PID:41513
GB:U00096 percent identity: 63.68; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="fumarase (fumC)"
/protein_id="AAD08368.1"
/db_xref="GI:2314492"
/translation="MQFRIEHTWGEIKVNDQYGAQTORSLRNFKIGTEKMPKELI
GAFAKRLAVNHRKLGKLSKQAIKACDILKGELOGEFFLAIWOTGSGTOTN
MNLNVIANKATEILGGNPREKKLHPNDQNMSSNDTFTAMHIVSVLTHITRLL
PSLENLTKTFKSGQKEIVKIGRTHIQDADPLTIGQFSGYASMLRSHKQOILLES
EHLRELAAGTAVGTGLNAHKELSEKVAELSOFGSVAPNPKPHALTSHDAIYA
HCAFALANLWKIANDIRWLASGPRCGLGELNIPENEPGSSIMPCKVNPOTCEAMTM
YAVQVMDNTAIGIASQGNFELNFKPVIYINFLQSLRLSDSNEHSHICASGIEP
NREKIDYILHSLMULVALNPHVGYENAAKIAKNAHKKGISLAKELKLLSAEDEFD
KPVVPEKMGPKA"
4515..4892
/gene="HP1326"
CDS
4515..4892
/gene="HP1326"
/notes="hypothetical protein; identified by GeneMark;
putative"
/codon_start=1
/transl_table=11
/product="H. pylori predicted coding region HP1326"
/protein_id="AAD08376.1"
/db_xref="GI:2314500"
/translation="MKKLAALFLVSLVGMVGLNABEQTLLKANDLEVKIKSVGNPKGD
NTFILSPILKGALEKAIYRVQFMPEMPGMPAMKMAQVSEKNGLYEAKTNLSMNGT
WQVRVDIKSEKQVYRAKTSLDL"
4897..6135
/gene="HP1327"
CDS
4897..6135
/gene="HP1327"
/notes="hypothetical protein; identified by GeneMark;
putative"
/codon_start=1
/transl_table=11
/product="H. pylori predicted coding region HP1327"
/protein_id="AAD08377.1"
/db_xref="GI:2314501"
/translation="MLSFISAFDKRGVSIRLTALLLLFSLGLAKDLEIQSFVAKYLS
KNQIQALQEQIDALSSQEKVKVDNPIILYLYNNANVDFRLDSTLMONKSLGLS
QVLDNGKRLQDSQMDLEKQKKILELTKTKQOLALSLMINGIENKKNQOELELTKTA
IKNLNTQIQAHHSSPNLIAIAKLEILKSQLEIKNNLEALSASHYSVMGELAFKEN
ELLSAPKNFENRQELNHSATNADIAIRLDEKSKOKDTLAKKSFLEDAVNTGV
YVFRSKQYNYDFEIALSIPLIYVQKQAKLVQEKESLSEKSEVNTKNTKTHIAL
KLLKLETLQKLNESINKIKQNEKIAQIAYLDLKSNGDYNNAYNAFNDKITQITQL
FTLSALNSTYLSLQNLKGL"
6132..7148
/gene="HP1328"
CDS
6132..7148
/gene="HP1328"
/notes="similar to GB:M26073 SP:P13510 PID:141927 percent
identity: 28.85; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11
/product="cation efflux system protein (czcA)"
/protein_id="AAD08369.1"
/db_xref="GI:2314493"
/translation="MKRIILWALILFFSLPFLANAQTOEIKKTEAKSOTRPNISTTK
VTEKFSOSRRYALLENPEALIFQTLRFDGYVEKLYANKYTIYTKKGRLLSYSP
VLEASQSLSEKFNQVGAIKERKILGLSSEKISSTIKSHVQNEHTIYSHFNGI
IPKSPDLNKGSIKQGBELFOIDLQWALVAVKYNQOELEFLKNTKAILPEVGIKG
EQEITLENINPINFKEDKMLEARENVPNVKQIYYPNMFAQVEIFQKPKMKILPREAV
```

```

LKGGKAIIVFKKDDFGLSPLEIKAVRLSDSGSYEILEGLKAGEEVANNALFVLADAAQN
NGDY"
7149..10256
/gene="HP1329"
CDS
7149..10256
/gene="HP1329"
/notes="similar to GB:M26073 SP:P13511 PID:141928 percent
identity: 31.32; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11
/product="cation efflux system protein (czcA)"
/protein_id="AAD08370.1"
/db_xref="GI:2314494"
/translation="MIEKIIDLVSNNKLLTTLVTLIFLFIASLWAKSVKRLDLPDLSP
AQVVOITVPNOSPQKIVQEQVTVPLVSTFMSIANTDVRGSISSYEGGLIYIIFKQGVN
LYWADRVLQENRVSNLFPKDAKVEIGSDSTIGWAYQYALSQKNSDLKVLQDFY
VYVALGTVGVSEVASGVGFVADYEVTQNDLSLIRYLNLSLEQVANAIKNSNDTGGV
ILENGFEKTIIRSHGYIQSLKDLKEIIVWKKEGAIPLKINDIASVRLTPKPRCAANLNG

Alignment Scores:
Pred. No.: 74.8 Length: 12496
Score: 57.00 Matches: 12
Percent Similarity: 69.57% Conservative: 4
Best Local Similarity: 52.17% Mismatches: 7
Query Match: 44.19% Indels: 0
DB: 1 Gaps: 0

US-09-847-539a-6_COPY_59_86 (1-28) x AE000634 (1-12496)

Qy 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLys 23
      :::::||||||| :::::||||||| :::::||||||| :::::|||||||
Db 5041 ATCAAGCCTTACAGAGCAATTTGACGCTTTAAGTTCTCAAGAAAAGTCGTTACCAAC 5100

Qy 24 AlaAspAsn 26
      |||||
Db 5101 TGGGATAAC 5109

RESULT 9
AE001548
LOCUS
DEFINITION
Helicobacter pylori, strain J99 section 109 of 132 of the complete
genome.
ACCESSION AE001548 AE001439
VERSION AE001548.1 GI:4155845
KEYWORDS
SOURCE
Helicobacter pylori J99.
ORGANISM
Helicobacter pylori J99
Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
Helicobacter.
REFERENCE
1 (bases 1 to 12526)
Alm,R.A., Ling,L.S., Moir,D.T., King,B.L., Brown,E.D., Dolg,P.C.,
Smith,D.R., Noonan,B., Guild,B.C., deJonge,B.L., Carmel,G.,
Tummino,P.J., Caruso,A., Uria-Nickelsen,M., Mills,D.M., Ives,C.,
Gibson,R., Merberg,D., Mills,S.D., Jiang,Q., Taylor,D.E.,
Vovis,G.F. and Trust,T.J.
Genomic-sequence comparison of two unrelated isolates of the human
gastric pathogen Helicobacter pylori
Nature 397 (6715), 176-180 (1999)
99120557
Erratum: [[published erratum appears in Nature 1999 Feb
25;397(6721):719]]
2 (bases 1 to 12526)
King,B.L., Alm,R.A. and Trust,T.J.
Direct Submission
Submitted (12-JAN-1999) Astra Research Center Boston, 128 Sidney
Street, Cambridge, MA 02139, USA
Address all correspondence to: hp@arch.us.astro.com or Richard
A. Alm, Astra Research Center Boston, 128 Sidney Street, Cambridge,
MA, 02139. Lo-See L. Ling, Donald T. Moir, Douglas R. Smith,
Braydon C. Guild, Gilles Carmel, Anthony Caruso, Debra M. Mills,
Rene Gibson, and Gerald F. Vovis are with Genome Therapeutics
Corporation, 100 Beaver Street, Waltham, MA, 02453. Qin Jiang and
```

sig\_peptide 665..733  
CDS /note="S antigen signal peptide"  
665..1594 /note="S antigen precursor"  
/codon\_start=1  
/protein\_id="AA29758.1"  
/db\_xref="GI:160871"  
/translation="MNRILSVFLLYLYIKYGVKNTDOEISNIYGTNYLRN  
GFLNGKNGKNKYEDLQEGEGENDEHSNDEENIIVGQDEAPKSDAEA  
LKDEALKSDAEARKSDEAEALKSDEAEARKSDEAEALKSDEAEARK  
SDEAEALKSDEAEARKSDEAEARKSDEAEARKSDEAEARKSDEAEALKSD  
EAEARKSDEAEARKSDEAEALKSDEAEARKSDEAEARKSDEAEARKSD  
EGPAGTGPGGSHNKKKSKSIMMLILM"  
734..1591  
/product="S antigen"  
BASE COUNT 1601 a 270 c 504 g 1463 t  
ORIGIN 1 bp upstream of EcoRI site.

Alignment Scores:  
Pred. No.: 17.9 Length: 3838  
Score: 57.50 Matches: 14  
Percent Similarity: 71.43% Conservative: 6  
Best Local Similarity: 50.00% Mismatches: 7  
Query Match: 44.57% Indels: 1  
DB: 3 Gaps: 1

US-09-847-539a-6\_COPY\_59\_86 (1-28) x PFASA7 (1-3838)

Qy 1 SerAspAlaLeuGluAlaLeu---AlaAspGlnThrAspAlaLeuGlnSerCluGluAla 19  
Db 953 AGTGATGAGGACGAGGCACTAAAGTGTATGAGGACGAGGCACTAAAGTGTATGAGGCA 1012  
Qy 20 AlavalValLyAlaAspAsnAla 27  
Db 1013 GAGGCACGAAAGTGTATGAGGCA 1036

RESULT 8  
AE000634  
LOCUS  
DEFINITION Helicobacter pylori 26695 section 112 of 134 of the complete genome.  
ACCESSION AE000634 AE000511  
VERSION AE000634.1 GI:2314489  
KEYWORDS  
SOURCE Helicobacter pylori 26695.  
ORGANISM Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group; Helicobacter.  
REFERENCE 1 (bases 1 to 12496)  
Tomb,J.-F., White,O., Kerlavage,A.R., Clayton,R.A., Sutton,G.G., Fleischmann,R.D., Ketchum,K.A., Klenk,H.P., Gill,S., Dougherty,B.A., Nelson,K., Quackenbush,J., Zhou,L., Kirkness,E.F., Peterson,S., Loftus,B., Richardson,D., Dodson,R., Khalak,H.G., Glodek,A., McKenney,K., Fitzegerald,L.M., Lee,N., Adams,M.D., Hickey,E.K., Berg,D.E., Gocayne,J.D., Utterback,T.R., Peterson,J.D., Kelley,J.M., Karp,P.D., Smith,H.O., Fraser,C.M. and Venter,J.C.  
The complete genome sequence of the gastric pathogen Helicobacter pylori  
J Nature 388 (6642), 539-547 (1997)  
JOURNAL  
MEDLINE 97394467  
REMARK Erratum:[published erratum appears in Nature 1997 Sep 25:389(6649):412]]  
REFERENCE 2 (bases 1 to 12496)  
Tomb,J.-F., White,O., Kerlavage,A.R., Clayton,R.A., Sutton,G.G., Fleischmann,R.D., Ketchum,K.A., Klenk,H.P., Gill,S., Dougherty,B.A., Nelson,K., Quackenbush,J., Zhou,L., Kirkness,E.F., Peterson,S., Loftus,B., Richardson,D., Dodson,R., Khalak,H.G., Glodek,A., McKenney,K., Fitzegerald,L.M., Lee,N., Adams,M.D., Hickey,E.K., Berg,D.E., Gocayne,J.D., Utterback,T.R., Peterson,J.D., Kelley,J.M., Cotton,M.D., Meldman,J.M., Fujii,C., Bowman,C., Watthey,L., Wallin,E., Hayes,W.S., Borodovsky,M., Karp,P.D., Smith,H.O., Fraser,C.M. and Venter,J.C.

TITLE Direct Submission  
JOURNAL Submitted (06-AUG-1997) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA  
REFERENCE 3 (bases 1 to 12496)  
AUTHORS White,O.  
TITLE Direct Submission  
JOURNAL Submitted (17-MAR-1999) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA  
LOCATION/Qualifiers  
FEATURES  
source 1..12496  
/organism="Helicobacter pylori 26695"  
/strain="26695"  
/db\_xref="taxon:85962"  
113..1246  
/gene="HP1321"  
113..1246  
/gene="HP1321"  
note="similar to GB:L77117 PID:L592175 percent identity: 30.81; identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical ATP-binding protein"  
/protein\_id="AAD08366.1"  
/db\_xref="GI:2314490"  
/translation="MRYDYLNPISFERKFWPLSCLHPGPFETPKPEVLPNPLLKAPS  
SDKICLIGPMKSGKTFALKLAKVKNPVIYINQMLNQNILSSWLLKWLKKMDL  
LILDRIDLPFLPKPLKIVLPNCLSPITAPNFSLCVAGLNFKEYTSFQKPTPKN  
TLNRFRLDGNALDSLTENOEKILKQENIOLIFQAYAPLMAKICGYSKFSYAFY  
LYTLKKELKSTDLTKLLHALEKQRIELVLPNFKNTKLYLCDFALPYSLTPSPS  
LLNVFENMFLELYKQPKYELYSHDNGIFILRENSTNKALIAHAHPTTFLEKQLL  
WCHKHGFNLIIIVSINAPISATNTPYKHLNFDISLDIQSLV"  
1452..2057  
/gene="HP1322"  
1452..2057  
/gene="HP1322"  
note="hypothetical protein; identified by GeneMark; putative"  
/codon\_start=1  
/transl\_table=11  
/product="H. pylori predicted coding region HP1322"  
/protein\_id="AAD08374.1"  
/db\_xref="GI:2314498"  
/translation="MIFTRKATWNALKLSFCALLSLGLFAQTAHAKHLKGTINYPD  
WLEINFFDEKNPPNQYGSASISGRNDFYANYIPYDDLPPEQNAEKIALRARINA  
YSTLESILLTKMHNRIYKVLQNNVISHLFGLVDFLTSKILAKRFVDVTTNHRVYV  
VQFFIOPEDLIAVFKAKRIDLSSASATHLSALLNKALFHL"  
complement(2031..2660)  
/gene="HP1323"  
complement(2031..2660)  
/gene="HP1323"  
note="similar to GB:L42023 SP:P43808 PID:1006316  
PID:1221175 PID:1205308 percent identity: 36.32;  
identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="ribonuclease HII (rnhB)"  
/protein\_id="AAD08367.1"  
/db\_xref="GI:2314491"  
/translation="MGCVSMITIGIDAGRGCLAGSLFVAGVACNEKTALEFLKMLKLD  
SKLSLAKRFFLEYKIKTHGEVGFVYVKKSSANEIDSLGLGACILKLAQVEILNCSLV  
DEIKIDNTAFGLNKRPIHTQITTINGDETIAQIAIASVLAFAKADREMLHLALFKEY  
GMDKNCYGTQKHIEAIIKLGATPFHRHSFTLKNRILNPKLLEVEQRLLI"  
complement(2676..2927)  
/gene="HP1324"  
complement(2676..2927)  
/gene="HP1324"  
note="hypothetical protein; identified by GeneMark; putative"  
/transl\_table=11  
/product="H. pylori predicted coding region HP1324"  
/protein\_id="AAD08375.1"  
/db\_xref="GI:2314499"





```
source      1. .832
/organism="Streptococcus pyogenes"
/strain="ATCC700294"
/db_xref="ATCC:700294"
/db_xref="taxon:1314"
sig_peptide 1. .99
CDS         1. .654
/feature="grab"
/codon_start=1
/transl_table=11
/product="GRAB precursor"
/protein_id="AAD26338.1"
/db_xref="GI:4589079"
/translation="MGKEIKVKFLRSAGFLVAVSALVGVSTVSAVDSPIEPRRI
PNGSTLNLGNAPKEKALRNEAIDELKKQATDEKATTAIEAASDALEALADOT
DALQSEAAVYKADNAASDALEALADOTDALQSEAEVYVSDNAASDAWEKAATPIAL
DVKKTKDKPVVKKKEERQNVNTLPTTGESNPFPTAAALALMWSTGLVSVSSCKEN"
mat_peptide 100. .651
/product="GRAB"
misc_feature 100. .273
/feature="Region: alpha2-macroglobulin-binding A domain"
misc_feature 274. .441
/feature="Region: repeat motif"
misc_feature 442. .576
/feature="Region: cell-wall attachment"
misc_feature 577. .651
/feature="Region: membrane-spanning"
BASE COUNT 304 a 143 c 169 g 216 t
ORIGIN

Alignment Scores:
Pred. No.:      1.4e-11      Length:      832
Score:          129.00      Matches:      28
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:      100.00%      Indels:      0
DB:              1          Gaps:      0

US-09-847-539a-6_COPY_59_86 (1-28) x AFL24399 (1-832)

Qy      1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAla 20
Db      274 TCAGATGCGCTTAGAGCATTACGGATCAACAGACGCTTTACAATCAGAGAAGCTGGC 333

Qy      21 ValVallyAlaAspAsnAlaAla 28
Db      334 GTTGTAAAGCGGATAACGCTGCT 357

RESULT 6
AE006573/c
LOCUS
DEFINITION Streptococcus pyogenes M1 GAS strain SF370, section 102 of 167 of
the complete genome.
ACCESSION AE006573 AE004092
VERSION AE006573.1 GI:13622459
KEYWORDS
SOURCE Streptococcus pyogenes M1 GAS.
ORGANISM Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
Streptococcus.
REFERENCE 1 (bases 1 to 10029)
Ferretti,J.J., McShan,W.M., Adjic,D., Savic,D., Savic,G., Lyon,K.,
Primeaux,C., Sezate,S.S., Surorov,A.N., Kenton,S., Lai,H., Lin,S.,
Qian,Y., Jia,H.G., Najjar,F.Z., Ren,Q., Zhu,H., Song,L., White,J.,
Yuan,X., Clifton,S.W., Roe,B.A. and McLaughlin,R.E.
Complete genome sequence of an M1 strain of Streptococcus pyogenes
Proc. Natl. Acad. Sci. U.S.A. 98 (8), 4658-4663 (2001)
21192684
MEDLINE 11296296
PUBMED
REFERENCE 2 (bases 1 to 10029)
Ferretti,J.J., McShan,W.M., Adjic,D., Savic,D., Savic,G., Lyon,K.,
Primeaux,C., Sezate,S.S., Surorov,A.N., Kenton,S., Lai,H., Lin,S.,
Qian,Y., Jia,H.G., Najjar,F.Z., Ren,Q., Zhu,H., Song,L., White,J.,
Yuan,X., Clifton,S.W., Roe,B.A. and McLaughlin,R.E.
Complete genome sequence of an M1 strain of Streptococcus pyogenes
Proc. Natl. Acad. Sci. U.S.A. 98 (8), 4658-4663 (2001)

Yuan,X., Clifton,S.W., Roe,B.A. and McLaughlin,R.E.
Direct Submission
Submitted (10-APR-2001) Department of Microbiology and Immunology,
University of Oklahoma Health Sciences Center, 940 SL Young Blvd,
Oklahoma
City, OK 73104, USA
Location/Qualifiers
1. .10029
/organism="Streptococcus pyogenes M1 GAS"
/strain="SF370"
/serotype="M1"
/db_xref="taxon:160490"
complement(135. .626)
/gene="arok"
/feature="SPy1351"
complement(135. .626)
/gene="arok"
/EC_number="2.7.1.71"
/feature="Best Blastp hit - Xylella fastidiosa (strain 9a5c)
XF1335 [imported] - Xylella fastidiosa"
>gi|19106332|gb|AA84144.1|AE003966_5 (AE003966) shikimate
kinase [Xylella fastidiosa]
/codon_start=1
/transl_table=11
/product="putative shikimate kinase"
/protein_id="AAK34179.1"
/db_xref="GI:13622460"
/translation="MTKVLGFMGVGKTTVSKHLSMHCKMDAITEAKIGMSIAAFFE
QHGRITAFRTIESOVKLDLFPANDNSIIVTGGVVVLOENROLLRNKHNNHLLVASPE
TLYORLKHDKKSQRPLFLKYSKEAFEFYQORMVFEGLSDLVIRVDHRTPEEVANII
EGY"
complement(619. .1911)
/gene="arok"
/feature="SPy1352"
complement(619. .1911)
/gene="arok"
/EC_number="2.5.1.19"
/feature="Best Blastp hit - sPlP43905|AROA_LACLA
3-PHOSPHOSHIMKIMATE 1-CARBOXYVINYLTRANSFERASE
(5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE) (EPSP
SYNTHASE) (EPSPS) >gi|1075724|pir||S52580
3-phosphoshikimate 1-carboxyvinyltransferase (EC 2.5.1.19)
- Lactococcus lactis >gi|683583|emb|CAA55180.1| (X78413)
5-enolpyruvylshikimate-3-phosphate synthase [Lactococcus
lactis]
/codon_start=1
/transl_table=11
/product="putative 3-phosphoshikimate
1-carboxyvinyltransferase"
/protein_id="AAK34180.1"
/db_xref="GI:13622461"
/translation="MKRMKLRNAGPLQGTIQVPGDKSISHRAVILGAVAKETRVRKG
LLGEDVLTIOAFNLGVRTVEKDDQVIEGQFQGLNAPCOTLNMGNSGTSMLRIA
LKGAQSPSVKMGDESKRPMORIVYPLKMGVEISGCTDRQFPPLQLOGNRLQIP
ITYLPISSAQVKSAILLAQAKQTTQVVEKETRNHTEMIQOFGRLVDKRRIT
LVGPOOLTAQETIVPGDISSAFWLVLGLIIPGSELLKUNGVNRTGLILEVVEKMG
AQIYEDNMKEQVTSIRVYSNMKGTIISGGLIPRLIDELPILATALLAQGTGCIK
DAQELRVKETRIOVVTIDILNSMGANIKATADGMIIKGTPLYGANTSTYGDHRIGMM
TAIALLVKQGVHLDKKEAJMTSYPTFFKDLERLCHD"
complement(2013. .2978)
/gene="SPy1353"
complement(2013. .2978)
/feature="SPy1353"
/feature="Best Blastp hit - dbj|BAB07074.1| (AP001518)
unknown conserved protein [Bacillus halodurans]"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK34181.1"
/db_xref="GI:13622462"
/translation="MAEKKWFDKVLKWKQYEPQVFMRLQSAEMDLISAIVAVYLIL
TAFPLIVTAANIIFYLNIADILLRLMKONLPKIDIFRPAISAIVENIFSKPSGVLGVA
TATGLWNTSRSLTSLOKAINKAYCASQRDFHVLGLLSLILFLALPALFSIF
SKAAIQVLDRKHYHLSDNITTIFLLLIQPIITVLLIFVGLMLLYFLPLPNVKIKRIRILP
```

REFERENCE 2 (bases 1 to 717)  
AUTHORS Rasmussen,M., Muller,H.P. and Bjorck,L.  
TITLE Direct Submission  
JOURNAL Submitted (28-JAN-1999) Cell and Molecular Biology, Molecular Pathogenesis, Soelvegatan 39, Lund 221 00, Sweden  
FEATURES  
source Location/Qualifiers  
1..717  
/organism="Streptococcus pyogenes"  
/strain="KT19"  
/db\_xref="taxon:1314"  
1..>717  
/product="GRAB"  
1..174  
/note="Region: alpha2-macroglobulin-binding A domain"  
1..>717  
/note="grab"  
/codon\_start=1  
/transl\_table=11  
/product="GRAB precursor"  
/protein\_id="AAD26341.1"  
/db\_xref="GI:4589085"  
/translation="VDSPIEQPRIIPNGGTLNLLGNAPKELALRNEERAIDELKKQA  
IEDKEATTAIEAASDALEADQADALQSEAAVQSDNAASDALEALADQDALQSE  
EAAVVKADNAASDTLEALADQTDALQSEAAVVKADNAASDTLEALADQDALQSE  
AAVVKADNAASDTLEALADQTDALQSEAAVQSDNAASDAWKAATPIALDVKKTD  
TKPVVKKERQNVNLTPTTGE"  
misc\_feature 175..594  
/note="Region: repeat motif"  
misc\_feature 595..717  
/note="Region: cell-wall attachment"  
BASE COUNT 257 a 140 c 163 g 157 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 1..19e-11 Length: 717  
Score: 129.00 Matches: 28  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 1 Gaps: 0  
US-09-847-539a-6\_COPY\_59\_86 (1-28) x AF124402 (1-717)  
QY 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluAlaAla 20  
|||||  
Db 259 AGTGACGCTTAGAAGCATGGCGGATCAACAGACGCTTTACAAATCAGAAGAGCTGCG 318  
QY 21 ValVallysAlaAspAsnAlaAla 28  
|||||  
Db 319 GTTGTAAAGCGGATACGCTGCT 342  
RESULT 4  
AF124403 804 bp DNA linear BCT 14-AUG-2000  
LOCUS Streptococcus pyogenes strain AP49 GRAB precursor, gene, partial  
DEFINITION cds.  
ACCESSION AF124403  
VERSION AF124403.1 GI:4589086  
KEYWORDS  
SOURCE Streptococcus pyogenes.  
ORGANISM Streptococcus pyogenes  
Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
Streptococcus.  
REFERENCE 1 (bases 1 to 804)  
AUTHORS Rasmussen,M., Muller,H.P. and Bjorck,L.  
TITLE Protein GRAB of streptococcus pyogenes regulates proteolysis at the bacterial surface by binding alpha2-macroglobulin  
J. Biol. Chem. 274 (22), 15336-15344 (1999)  
JOURNAL MEDLINE 99269061  
PUBMED 10336419  
REFERENCE 2 (bases 1 to 804)  
AUTHORS Rasmussen,M., Muller,H.P. and Bjorck,L.  
TITLE Direct Submission  
JOURNAL Submitted (28-JAN-1999) Cell and Molecular Biology, Molecular Pathogenesis, Soelvegatan 39, Lund 221 00, Sweden

JOURNAL Submitted (28-JAN-1999) Cell and Molecular Biology, Molecular Pathogenesis, Soelvegatan 39, Lund 221 00, Sweden  
FEATURES  
source Location/Qualifiers  
1..804  
/organism="Streptococcus pyogenes"  
/strain="AP49"  
/db\_xref="taxon:1314"  
1..>804  
/product="GRAB"  
1..174  
/note="Region: alpha2-macroglobulin-binding A domain"  
1..>804  
/note="grab"  
/codon\_start=1  
/transl\_table=11  
/product="GRAB precursor"  
/protein\_id="AAD26342.1"  
/db\_xref="GI:4589087"  
/translation="VDSPIEQPRIIPNGGTLNLLGNAPKELALRNEERAIDELKKQA  
IEDKEATTAIEAASDALEADQADALQSEAAVQSDNAASDALEALADQDALQSE  
EAAVVKADNAASDTLEALADQTDALQSEAAVVKADNAASDALEALADQDALQSE  
VQSDNAASDAWKAATPIALDVKKTKDTKPVVKKERQNVNLTPTTGE"  
misc\_feature 175..678  
/note="Region: repeat motif"  
misc\_feature 679..804  
/note="Region: cell-wall attachment"  
BASE COUNT 284 a 161 c 185 g 174 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 1..35e-11 Length: 804  
Score: 129.00 Matches: 28  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 1 Gaps: 0  
US-09-847-539a-6\_COPY\_59\_86 (1-28) x AF124403 (1-804)  
QY 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluAlaAla 20  
|||||  
Db 511 AGTGACGCTTAGAAGCATGGCGGATCAACAGACGCTTTACAAATCAGAAGAGCTGCG 570  
QY 21 ValVallysAlaAspAsnAlaAla 28  
|||||  
Db 571 GTTGTAAAGCGGATACGCTGCT 594  
RESULT 5  
AF124399 832 bp DNA linear BCT 14-AUG-2000  
LOCUS Streptococcus pyogenes strain ATCC700294 GRAB precursor, gene, complete cds.  
DEFINITION AF124399  
ACCESSION AF124399.1 GI:4589078  
VERSION AF124399.1  
KEYWORDS  
SOURCE Streptococcus pyogenes.  
ORGANISM Streptococcus pyogenes  
Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
Streptococcus.  
REFERENCE 1 (bases 1 to 832)  
AUTHORS Rasmussen,M., Muller,H.P. and Bjorck,L.  
TITLE Protein GRAB of streptococcus pyogenes regulates proteolysis at the bacterial surface by binding alpha2-macroglobulin  
J. Biol. Chem. 274 (22), 15336-15344 (1999)  
JOURNAL MEDLINE 99269061  
PUBMED 10336419  
REFERENCE 2 (bases 1 to 832)  
AUTHORS Rasmussen,M., Muller,H.P. and Bjorck,L.  
TITLE Direct Submission  
JOURNAL Submitted (28-JAN-1999) Cell and Molecular Biology, Molecular Pathogenesis, Soelvegatan 39, Lund 221 00, Sweden  
FEATURES  
Location/Qualifiers

```

ORGANISM Streptococcus pyogenes
Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
Streptococcus.
REFERENCE 1 (bases 1 to 468)
AUTHORS Rasmussen,M., Muller,H.P. and Bjorck,L.
TITLE Protein GRAB of streptococcus pyogenes regulates proteolysis at the
bacterial surface by binding alpha2-macroglobulin
J. Biol. Chem. 274 (22), 15336-15344 (1999)
JOURNAL
MEDLINE 99269061
PUBMED 10336419
REFERENCE 2 (bases 1 to 468)
AUTHORS Rasmussen,M., Muller,H.P. and Bjorck,L.
TITLE Direct Submission
JOURNAL Submitted (28-JAN-1999) Cell and Molecular Biology, Molecular
Pathogenesis, Soelvegatan 39, Lund 221 00, Sweden
FEATURES
source
1..468
Location/Qualifiers
/organism="Streptococcus pyogenes"
/strain="KTI3"
/db_xref="taxon:1314"
/product="GRAB"
/mat_peptide 1..>468
misc_feature 1..174
CDS
<1..>468
/note="Region: alpha2-macroglobulin-binding A domain"
/translation="VDSPIEQRIIPNGTTLNLGNAPKIALRNEPRAIDELKKQA
IEDKEATTAEASDSDALEADQTDALQSEAAVVKADNAASDALEALADQTDALQS
EEAEVQSDNAASDAWEKAATPIALDVKKTKDKPVVKKERQNVNTLPTTCEE"
misc_feature 175..342
misc_feature 343..468
/note="Region: repeat motif"
BASE COUNT 178 a 91 c 100 g 99 t
ORIGIN
Alignment Scores:
Pred. No.: 7 61e-12 Length: 468
Score: 129.00 Matches: 28
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 1 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x AF124401 (1-468)

QY 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAla 20
Db 175 TCAGATGCCTTAGAAGCATTAGCGGATCAACAGACGCTTTACAATCAGAAGAAGCTGCG 234

QY 21 ValValLysAlaAspAsnAlaAla 28
Db 235 GTTGTAAAGCGGATAACGCTGCT 258

RESULT 2
AF124400
LOCUS AF124400 469 bp DNA linear BCT 14-AUG-2000
DEFINITION Streptococcus pyogenes strain AP1 GRAB precursor, gene, partial
cfs.
ACCESSION AF124400
VERSION AF124400.1 GI:4589080
KEYWORDS Streptococcus pyogenes.
SOURCE Streptococcus pyogenes.
ORGANISM Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
Streptococcus.
REFERENCE 1 (bases 1 to 469)
AUTHORS Rasmussen,M., Muller,H.P. and Bjorck,L.
TITLE Protein GRAB of streptococcus pyogenes regulates proteolysis at the
bacterial surface by binding alpha2-macroglobulin
J. Biol. Chem. 274 (22), 15336-15344 (1999)
JOURNAL
MEDLINE 99269061
PUBMED 10336419

```

```

TITLE Protein GRAB of streptococcus pyogenes regulates proteolysis at the
bacterial surface by binding alpha2-macroglobulin
J. Biol. Chem. 274 (22), 15336-15344 (1999)
MEDLINE 99269061
PUBMED 10336419
REFERENCE 2 (bases 1 to 469)
AUTHORS Rasmussen,M., Muller,H.P. and Bjorck,L.
TITLE Direct Submission
JOURNAL Submitted (28-JAN-1999) Cell and Molecular Biology, Molecular
Pathogenesis, Soelvegatan 39, Lund 221 00, Sweden
FEATURES
source
1..469
Location/Qualifiers
/organism="Streptococcus pyogenes"
/strain="AP1"
/db_xref="taxon:1314"
/product="GRAB"
/mat_peptide 1..>469
misc_feature 1..174
CDS
<1..>469
/note="Region: alpha2-macroglobulin-binding A domain"
/translation="VDSPIEQRIIPNGTTLNLGNAPKIALRNEPRAIDELKKQA
IEDKEATTAEASDSDALEADQTDALQSEAAVVKADNAASDALEALADQTDALQS
EEAEVQSDNAASDAWEKAATPIALDVKKTKDKPVVKKERQNVNTLPTTCEE"
misc_feature 175..342
misc_feature 343..469
/note="Region: repeat motif"
BASE COUNT 178 a 90 c 100 g 101 t
ORIGIN
Alignment Scores:
Pred. No.: 7 63e-12 Length: 469
Score: 129.00 Matches: 28
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 1 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x AF124400 (1-469)

QY 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAla 20
Db 175 TCAGATGCCTTAGAAGCATTAGCGGATCAACAGACGCTTTACAATCAGAAGAAGCTGCG 234

QY 21 ValValLysAlaAspAsnAlaAla 28
Db 235 GTTGTAAAGCGGATAACGCTGCT 258

RESULT 3
AF124402
LOCUS AF124402 717 bp DNA linear BCT 14-AUG-2000
DEFINITION Streptococcus pyogenes strain KTL9 GRAB precursor, gene, partial
cfs.
ACCESSION AF124402
VERSION AF124402.1 GI:4589084
KEYWORDS Streptococcus pyogenes.
SOURCE Streptococcus pyogenes.
ORGANISM Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
Streptococcus.
REFERENCE 1 (bases 1 to 717)
AUTHORS Rasmussen,M., Muller,H.P. and Bjorck,L.
TITLE Protein GRAB of streptococcus pyogenes regulates proteolysis at the
bacterial surface by binding alpha2-macroglobulin
J. Biol. Chem. 274 (22), 15336-15344 (1999)
JOURNAL
MEDLINE 99269061
PUBMED 10336419

```

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 13, 2002, 04:49:23 ; Search time 274.909 Seconds  
(without alignments)  
2131.407 Million cell updates/sec

Title: US-09-847-539A-6\_COPY\_59\_86

Perfect score: 129

Sequence: 1 SDALEALADQTDALQSEAAVVKADNAA 28

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame\_p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool/US09847539/runat\_10102002\_093105\_4992/app\_query.fasta\_1.526  
-DB=GenEmbl -QFMT=fastap -SUFFIX=std.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40 cdi -LIST=45  
-DOCALIGN=200 -THR\_SCORE=500 -THR\_MIN=0 -THR\_MAX=100 -ALIGN=40 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEA=SIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09847539.cgn\_1\_1182\_@runat\_10102002\_093105\_4992 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.\*

- 1: gb\_ba.\*
- 2: gb\_htg.\*
- 3: gb\_in.\*
- 4: gb\_cm.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vi.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_mu.\*
- 20: em\_or.\*
- 21: em\_ov.\*
- 22: em\_pat.\*
- 23: em\_ph.\*
- 24: em\_pi.\*
- 25: em\_ro.\*
- 26: em\_sts.\*
- 27: em\_un.\*
- 28: em\_un.\*

29: em\_vi.\*  
30: em\_htg\_hum.\*  
31: em\_htg\_inv.\*  
32: em\_htg\_other.\*  
33: em\_htgo\_inv.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	129	100.0	468	1	AF124401 Streptoco
2	129	100.0	469	1	AF124400 Streptoco
3	129	100.0	717	1	AF124402 Streptoco
4	129	100.0	804	1	AF124403 Streptoco
5	129	100.0	832	1	AF124399 Streptoco
6	129	100.0	10029	1	AE006573 Streptoco
7	57.5	44.6	3838	3	PFASA7
8	57	44.2	12496	1	AE000634 Helicobac
9	57	44.2	12526	1	AE001548 Helicobac
10	57	44.2	106516	1	STYPPHCM2
11	54	41.9	1288	3	AF031321 Tetrahyme
12	54	41.9	32270	2	AC014470 Drosophil
13	54	41.9	151610	3	AC009462 Drosophil
14	54	41.9	155840	2	AC006495 Drosophil
15	54	41.9	224896	3	AE003721 Drosophil
16	53	41.1	7686	1	SC0414671 Streptomy
17	53	41.1	31317	1	SCSCP2
18	53	41.1	117309	9	AC007274 Homo sapi
19	53	41.1	173772	2	AC087862 Homo sapi
20	52.5	40.7	12411	1	AE004659 Pseudomon
21	52	40.3	12373	3	DME17920 Drosophila
22	52	40.3	16135	3	AF096897 Drosophil
23	52	40.3	17595	3	AF096896 Drosophil
24	52	40.3	36993	2	AC018007 Drosophil
25	52	40.3	133155	2	AC011910 Drosophil
26	52	40.3	169427	3	AC007416 Drosophil
27	52	40.3	281993	3	AE003619 Drosophil
28	52	40.3	303100	1	RME603646 Rhizobium
29	51.5	39.9	850	33	AC056141 Giardia
30	51.5	39.9	2888	1	ATU1522A Agrobacteri
31	51.5	39.9	12976	3	AC001654 Drosophil
32	51.5	39.9	32608	1	SCBAC36F5 Streptomy
33	51.5	39.9	35031	2	AC012724 Drosophil
34	51.5	39.9	110000	3	Continuation (3 of
35	51.5	39.9	170801	3	AC095014 Drosophil
36	51.5	39.9	190642	3	AC095015 Drosophil
37	51.5	39.9	309357	3	AE003673 Drosophil
38	51	39.5	2151	3	AY061146 Drosophil
39	51	39.5	12423	1	AE005155 Halobacte
40	51	39.5	14941	3	AF072878 Tetrahyme
41	51	39.5	20217	1	AE000047 Mycoplasma
42	51	39.5	30561	1	SC1F2 Streptomy
43	51	39.5	96288	3	AC005749 Drosophil
44	51	39.5	105823	2	AC019871 Drosophil
45	51	39.5	135509	8	AC037426 Oryza sat

ALIGNMENTS

RESULT 1

AF124401

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

AF124401 Streptococcus pyogenes strain K1L3 GRAB precursor, gene, partial cds.  
AF124401.1 GI:4589082  
Streptococcus pyogenes.

XX The present invention provides a number of single exon nucleic acid  
CC probes which are derived from genomic sequences expressed in the human  
CC bone marrow. They can be used to measure gene expression in bone marrow  
CC samples, which may enable the improved diagnosis and treatment of cancers  
CC such as lymphoma, leukaemia and myeloma. The present sequence is one of  
CC the probes of the invention.

XX  
SQ Sequence 544 BP; 124 A; 113 C; 135 G; 172 T; 0 other;

Alignment Scores:  
Pred. No.: 85.9 Length: 544  
Score: 45.50 Matches: 13  
Percent Similarity: 55.56% Conservative: 2  
Best Local Similarity: 48.15% Mismatches: 7  
Query Match: 35.27% Indels: 5  
DB: 22 Gaps: 1

US-09-847-539a-6\_COPY\_59\_86 (1-28) x AAK36405 (1-544)

Qy 2 AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSer----- 16  
Db 290 GACATTTTGGACAGACTGCAGCTCAGGTGGATGCCCTTGCAGGGCCATCTTCAACAAATG 349  
Qy 17 GluGluAlaAlaValValLys 23  
Db 350 GAGCAGGAAGCCCTGTGTGAAG 370

Search completed: October 13, 2002, 05:18:52  
Job time : 1037.53 secs

```
XX DR WPI; 2001-483447/52.
XX DR Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human fetal liver -
XX PS
XX PS Claim 1; SEQ ID NO 10489; 639pp + sequence listing; English.
XX CC
XX CC The invention relates to a single exon nucleic acid probe for
CC measuring human gene expression in a sample derived from human foetal
CC liver. The single exon nucleic acid probes may be used for predicting,
CC measuring and displaying gene expression in samples derived from human
CC fetal liver. The present sequence is a single exon nucleic acid
CC probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 544 BP; 124 A; 113 C; 135 G; 172 T; 0 other;
XX
Alignment Scores:      85.9      Length:      544
Pred. No.:      45.50      Matches:      13
Score:      55.56%      Conservative:      2
Best Local Similarity:      48.15%      Mismatches:      7
Query Match:      35.27%      Indels:      5
DB:      22      Gaps:      1
US-09-847-539A-6_COPY_59_86 (1-28) x ABA62184 (1-544)
QY 2 AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSer----- 16
Db 290 GACATTTTGGAAACAGACTGCAGCTGCGTGGATGCTTGCAGGGCCATCTTCACAAATG 349
QY 17 GluGluAlaAlaValValLys 23
Db 350 GAGCAGGAAGCCCTGGTGAAG 370
RESULT 39
AAK10508
ID AAK10508 standard; DNA; 544 BP.
XX AC AAK10508;
XX DT
XX DE 05-NOV-2001 (first entry)
XX DE Human brain expressed single exon probe SEQ ID NO: 10499.
XX KW Human; brain expressed exon; gene expression analysis; probe;
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW epilepsy; cancer; ss.
XX OS Homo sapiens.
XX OS WO200157275-A2.
XX PN 09-AUG-2001.
XX PD 30-JAN-2001; 2001WO-US00667.
XX PF 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483446/52.
XX DR
```

```
XX PT Single exon nucleic acid probes for analyzing gene expression in human
XX brains -
XX PS
XX PS Example 4; SEQ ID NO: 10499; 650pp + Sequence Listing; English.
XX CC
XX CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is one of the probes of the
CC invention.
XX SQ Sequence 544 BP; 124 A; 113 C; 135 G; 172 T; 0 other;
XX
Alignment Scores:      85.9      Length:      544
Pred. No.:      45.50      Matches:      13
Score:      55.56%      Conservative:      2
Best Local Similarity:      48.15%      Mismatches:      7
Query Match:      35.27%      Indels:      5
DB:      22      Gaps:      1
US-09-847-539A-6_COPY_59_86 (1-28) x AAK10508 (1-544)
QY 2 AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSer----- 16
Db 290 GACATTTTGGAAACAGACTGCAGCTGCGTGGATGCTTGCAGGGCCATCTTCACAAATG 349
QY 17 GluGluAlaAlaValValLys 23
Db 350 GAGCAGGAAGCCCTGGTGAAG 370
RESULT 40
AAK36405
ID AAK36405 standard; DNA; 544 BP.
XX AC AAK36405;
XX DT
XX DE 06-NOV-2001 (first entry)
XX DE Human bone marrow expressed single exon probe SEQ ID NO: 10962.
XX KW Human; bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX OS Homo sapiens.
XX OS WO200157276-A2.
XX PN 09-AUG-2001.
XX PD 30-JAN-2001; 2001WO-US00668.
XX PF 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488900/53.
XX DR
XX DR Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human bone marrow -
XX PS
XX PS Example 4; SEQ ID NO: 10962; 658pp + Sequence Listing; English.
XX PS
```

```
XX PF 30-JAN-2001; 2001WO-US00668.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX DR WPI; 2001-488900/53.
XX XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human bone marrow -
XX XX
XX PS Example 4; SEQ ID NO: 23882; 658pp + Sequence Listing; English.
XX CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukemia and myeloma. The present sequence is one of
CC the probes of the invention.
XX SQ Sequence 234 BP; 61 A; 55 C; 61 G; 57 T; 0 other;
Alignment Scores:
Pred. No.: 29.6 Length: 234
Score: 45.50 Matches: 13
Percent Similarity: 55.56% Conservative: 2
Best Local Similarity: 48.15% Mismatches: 7
Query Match: 35.27% Indels: 5
DB: 22 Gaps: 1
US-09-847-539A-6_COPY_59_86 (1-28) x AAK49325 (1-234)
Qy 2 AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSer----- 16
Db 154 GACATTTTGGACAGACTGCAGCTCAGGTGGATGCTTGCAGGGCCATCTTCAACAAATG 213
Qy 17 GluGluAlaAlaValVallys 23
Db 214 GAGCAGGAAGCCCTGGTGAAG 234
RESULT 37
AAI55172
ID AAI55172 standard; DNA; 234 BP.
XX AC AAI55172;
XX DT 17-OCT-2001 (first entry)
XX DE
XX DE Probe #23858 used to measure gene expression in human placenta sample.
XX KW Probe; microarray; human; placenta; antenatal diagnosis;
XX KW genetic disorder; ss.
XX OS Homo sapiens.
XX PN WO200157272-A2.
XX XX
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00663.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00663.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
```

```
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX DR WPI; 2001-48897/53.
XX XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human placenta -
XX XX
XX PS Claim 25; SEQ ID NO 23858; 654pp; English.
XX CC The present invention relates to single exon nucleic acid probes (SENP).
CC The present sequence is one such probe. The probes are useful for
CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.
XX SQ Sequence 234 BP; 61 A; 55 C; 61 G; 57 T; 0 other;
Alignment Scores:
Pred. No.: 29.6 Length: 234
Score: 45.50 Matches: 13
Percent Similarity: 55.56% Conservative: 2
Best Local Similarity: 48.15% Mismatches: 7
Query Match: 35.27% Indels: 5
DB: 22 Gaps: 1
US-09-847-539A-6_COPY_59_86 (1-28) x AAI55172 (1-234)
Qy 2 AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSer----- 16
Db 154 GACATTTTGGACAGACTGCAGCTCAGGTGGATGCTTGCAGGGCCATCTTCAACAAATG 213
Qy 17 GluGluAlaAlaValVallys 23
Db 214 GAGCAGGAAGCCCTGGTGAAG 234
RESULT 38
ABA62184
ID ABA62184 standard; DNA; 544 BP.
XX AC ABA62184;
XX DT 01-FEB-2002 (first entry)
XX DE
XX DE Human foetal liver single exon nucleic acid probe #10489.
XX KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX OS Homo sapiens.
XX PN WO200157277-A2.
XX XX
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00669.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
```

KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.

XX Homo sapiens.

OS WO200157277-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00669.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234587.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-483447/52.

XX Human genome-derived single exon nucleic acid probes useful for

XX analyzing gene expression in human fetal liver.

XX Claim 4; SEQ ID NO 22983; 639pp + sequence listing; English.

XX The invention relates to a single exon nucleic acid probe for  
CC measuring human gene expression in a sample derived from human foetal  
CC liver. The single exon nucleic acid probes may be used for predicting,  
CC measuring and displaying gene expression in samples derived from human  
CC foetal liver. The present sequence is a single exon nucleic acid  
CC probe of the invention.

CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 234 BP; 61 A; 55 C; 61 G; 57 T; 0 other;

Alignment Scores:

Pred. No.: 29.6 Length: 234  
Score: 45.50 Matches: 13  
Percent Similarity: 55.56% Conservative: 2  
Best Local Similarity: 48.15% Mismatches: 7  
Query Match: 35.27% Indels: 5  
DB: 22 Gaps: 1

US-09-847-539A-6\_COPY\_59\_86 (1-28) x ABA74678 (1-234)

OY 2 AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSer----- 16

DB 154 GACATTGGAACAGACTGCAGCTCAGGTGGATGCCTTGCAGGGCCATCTTCAACAAATG 213

OY 17 GluGluAlaAlaValVallys 23

DB 214 GAGCAGGAAGCCCTGGTGAAG 234

RESULT 35

AAK23154

ID AAK23154 standard; DNA; 234 BP.

XX AAK23154;

XX 05-NOV-2001 (first entry)

XX Human brain expressed single exon probe SEQ ID NO: 23145.

XX Human; brain expressed exon; gene expression analysis; probe;

XX microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;

XX epilepsy; cancer; ss.

OS Homo sapiens.

XX WO200157275-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00667.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234587.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-483446/52.

XX Single exon nucleic acid probes for analyzing gene expression in human

XX brains.

XX Example 4; SEQ ID NO: 23145; 650pp + Sequence Listing; English.

XX The present invention provides a number of single exon nucleic acid  
CC probes which are derived from genomic sequences expressed in the human  
CC brain. They can be used to measure gene expression in brain cell samples,  
CC which may enable the diagnosis and improved treatment of nervous system  
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,  
CC epilepsy and cancers. The present sequence is one of the probes of the  
CC invention.

XX Sequence 234 BP; 61 A; 55 C; 61 G; 57 T; 0 other;

Alignment Scores:

Pred. No.: 29.6 Length: 234  
Score: 45.50 Matches: 13  
Percent Similarity: 55.56% Conservative: 2  
Best Local Similarity: 48.15% Mismatches: 7  
Query Match: 35.27% Indels: 5  
DB: 22 Gaps: 1

US-09-847-539A-6\_COPY\_59\_86 (1-28) x AAK23154 (1-234)

OY 2 AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSer----- 16

DB 154 GACATTGGAACAGACTGCAGCTCAGGTGGATGCCTTGCAGGGCCATCTTCAACAAATG 213

OY 17 GluGluAlaAlaValVallys 23

DB 214 GAGCAGGAAGCCCTGGTGAAG 234

RESULT 36

AAK49325

ID AAK49325 standard; DNA; 234 BP.

XX AAK49325;

XX 06-NOV-2001 (first entry)

XX Human bone marrow expressed single exon probe SEQ ID NO: 23882.

XX Human; bone marrow expressed exon; gene expression analysis; probe;

XX microarray; cancer; leukaemia; lymphoma; myeloma; ss.

XX Homo sapiens.

XX WO200157276-A2.

XX 09-AUG-2001.



```
PR 07-APR-2000; 2000JP-0159162.
PR 03-AUG-2000; 2000JP-0280988.
XX
PA (KYOW ) KYOWA HAKKO KOGYO KK.
XX
PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX
DR WPI; 2001-376931/40.
XX
XX Novel polynucleotides derived from Coryneform bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene
XX
XX Disclosure; SEQ ID NO: 7060; 246pp + Sequence Listing; English.
XX
XX The present invention provides a number of nucleotide and protein
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of coryneform bacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from coryneform bacterium, and identifying a homologue of a gene derived
CC from coryneform bacterium. Coryneform bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a nucleic acid described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
XX
SQ Sequence 349980 BP; 80289 A; 91081 C; 97378 G; 81232 T; 0 other;

Alignment Scores:
Pred. No.: 2.41e+05 Length: 349980
Score: 46.00 Matches: 9
Percent Similarity: 62.50% Conservative: 6
Best Local Similarity: 37.50% Mismatches: 9
Query Match: 35.66% Indels: 0
DB: 22 Gaps: 0

US-09-847-539a-6_COPY_59_86 (1-28) x AAH68525 (1-349980)

QY 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLys 23
Db 225435 ATTGAGCCTTACGCACCTGAGGACGACGAGCTTGAAGAGGGCGAGTCCGCAATCTACAAG 225494
QY 24 AlaAspAsnAla 27
Db 225495 TTTCGAGAGCGCT 225506

RESULT 33
AAAB1490
ID AAAB1490 standard; DNA; 1437668 BP.
XX
AC AAAB1490;
XX
XX 04-DEC-2000 (first entry)
DE
DE N. meningitidis B full length genome DNA sequence SEQ ID NO:1068.
XX
XX Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;
KW antigen; vaccine; diagnosis; infection; antibacterial; identification;
KW Meningococcus B; MenB; ds.
XX
OS Neisseria meningitidis.
XX
PN WO200022430-A2.
XX
XX 20-APR-2000.
PD
XX
XX 08-OCT-1999; 99WO-US23573.
PF
XX

09-OCT-1998; 98US-0103794.
30-APR-1999; 99US-0132068.
XX
PA (CHIR ) CHIRON CORP.
XX
PI Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;
PI Masignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;
PI Rappuoli R, Pizza M;
XX
DR WPI; 2000-318079/27.
XX
XX Isolated nucleotide sequences of Neisseria meningitidis which can be
PT used in the diagnosis and treatment of N. meningitidis infection and
PT other Neisserial infections, for example, N.gonorrhoea -
XX
XX Claim 7; Page 866-1272; 1760pp; English.
XX
XX The present invention describes methods of obtaining immunogenic
CC proteins from Neisseria genomic sequences. AAA81453 to AAA82414
CC represent specifically claimed Neisseria meningitidis genomic DNA
CC sequences; AAA81260 to AAA81303 and AAB25620 to AAB25663 represent
CC Neisseria DNA sequences and their corresponding proteins; AAA81254 to
CC AA81259 and AAA81304 to AAA81321 represent PCR primers used in the
CC isolation of Neisseria meningitidis DNA sequences; and AAA81322 to
CC AAA81452 represent Neisseria meningitidis MenB polynucleotide ORF
CC sequences, which are all used in the exemplification of the present
CC invention. The nucleic acid sequences, protein sequences, and antibodies
CC against them, can be used in the manufacture of a composition. The
CC composition can be used as a medicament (or in the manufacture of a
CC medicament) for treating, preventing or diagnosing infection due to
CC Neisserial bacteria. For example, some of the identified proteins could
CC be components of vaccines against Meningococcus B; against all serotypes;
CC and/or against all pathogenic Neisseriae. Identification of sequences
CC from the bacterium will also facilitate production of biological probes,
CC particularly organism-specific probes. Attempts to make efficacious
CC Meningococcus B vaccines have failed mainly due to antigen tolerance.
CC Multivalent vaccines have also been tried but none have successfully
CC overcome antigenic variability. The provision of further, complete
CC sequences may provide an opportunity to identify secreted or surface
CC exposed proteins that may be presumed targets for the immune system and
CC which are not antigenically variable or at least more conserved than
CC other more variable regions.
XX
SQ Sequence 1437668 BP; 344338 A; 353206 C; 385074 G; 355045 T; 5 other;

Alignment Scores:
Pred. No.: 1.21e+06 Length: 1437668
Score: 46.00 Matches: 11
Percent Similarity: 65.38% Conservative: 6
Best Local Similarity: 42.31% Mismatches: 9
Query Match: 35.66% Indels: 0
DB: 21 Gaps: 0

US-09-847-539a-6_COPY_59_86 (1-28) x AAA81490 (1-1437668)

QY 3 AlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVal 22
Db 411891 GCTTTGGAAGCGGTTTGAAAACTCGAAGCGCGCAAAATCCCGACGGCAGCATTTG 411950
QY 23 LysAlaAspAsnAlaAla 28
Db 411951 GAAGCGGCTGAAGCGCT 411968

RESULT 34
ABA74678
ID ABA74678 standard; DNA; 234 BP.
XX
AC ABA74678;
XX
XX 01-FEB-2002 (first entry)
DT
XX
XX Human foetal liver single exon nucleic acid probe #22983.
DE
XX
```

CC AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the  
 CC isolation of *Neisseria meningitidis* DNA sequences; and AAA81322 to  
 CC AAA81452 represent *Neisseria meningitidis* MenB polynucleotide ORF  
 CC sequences, which are all used in the exemplification of the present  
 CC invention. The nucleic acid sequences, protein sequences, and antibodies  
 CC against them, can be used in the manufacture of a composition. The  
 CC composition can be used as a medicament (or in the manufacture of a  
 CC medicament) for treating, preventing or diagnosing infection due to  
 CC *Neisseria meningitidis*. For example, some of the identified proteins could  
 CC be components of vaccines against *Neisseria meningitidis*; against all serotypes;  
 CC and/or against all pathogenic *Neisseria*. Identification of sequences  
 CC from the bacterium will also facilitate production of biological probes,  
 CC particularly organism-specific probes. Attempts to make efficacious  
 CC *Neisseria meningitidis* vaccines have failed mainly due to antigen tolerance.  
 CC Multivalent vaccines have also been tried but none have successfully  
 CC overcome antigenic variability. The provision of further, complete  
 CC sequences may provide an opportunity to identify secreted or surface  
 CC exposed proteins that may be presumed targets for the immune system and  
 CC which are not antigenically variable or at least more conserved than  
 CC other more variable regions.

XX Sequence 69936 BP; 16512 A; 19259 C; 17399 G; 16763 T; 3 other;

Alignment Scores:  
 Pred. No.: 3.27e+04 Length: 69936  
 Score: 46.00 Matches: 11  
 Percent Similarity: 65.38% Conservatives: 6  
 Best Local Similarity: 42.31% Mismatches: 9  
 Query Match: 35.66% Indels: 0  
 DB: 21 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x AAA81479 (1-69936)

QY 3 AlaleuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaVal 22  
 DB 69296 GCTTTGAAGCGCGTTTGGAAACTCGAAGCGCGCAAAATCCGAGCGGCAGCATG 69237  
 QY 23 LysAlaAspAsnAlaAla 28  
 DB 69236 GAAGCGGCTGAAGCGGCT 69219

RESULT 31

AAF21607  
 ID AAF21607 standard; DNA: 349980 BP.

AC AAF21607;

XX 13-MAR-2001 (first entry)

XX *Neisseria meningitidis* B nucleotide sequence SEQ ID NO:108.

XX *Neisseria meningitidis*; *Neisseria gonorrhoeae*; immunogenic; vaccine;  
 KW diagnosis; antigen; detection; infection; gene therapy; antibacterial;  
 KW ds.

XX *Neisseria meningitidis*.

XX WO200066791-A1.

XX 09-NOV-2000.

XX 08-MAR-2000; 2000WO-US05928.

XX 30-APR-1999; 99US-0132068.

XX 08-OCT-1999; 99WO-US23573.

XX 28-FEB-2000; 2000GB-0004695.

XX {CHIR } CHIRON CORP.

XX {GENO-} INST GENOMIC RES.

XX Pizza M, Hickey E, Peterson J, Tettelin H, Venter JC, Maignani V;  
 PI Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V, Rappuoli R;  
 PI Frazer CM, Grandi G;

XX

XX WPI; 2000-647603/62.

XX *Neisseria meningitidis* B full length genome sequence and open reading  
 PT frames are used to detect, treat and prevent *Neisseria* infections -  
 XX

PS Claim 7; Appendix A; 692pp; English.

XX

CC The present invention describes the full length genome of  
 CC *Neisseria meningitidis* B (NMB). The sequences in AAF21544 and AAF21607  
 CC to AAF21613 represent fragments of the NMB genomic sequence, as the  
 CC sequence was too long to go in a record on its own it was split into 8  
 CC sequences which overlap each other at the beginning and end of each  
 CC sequence by 49980 bp (i.e. the last 49980 bp of AAF21544 is repeated at  
 CC the beginning of AAF21607, the last 49980 bp of AAF21607 are repeated at  
 CC the beginning of AAF21608, and so on). AAF21545 to AAF21588 encode the  
 CC *Neisseria* proteins given in AAF58550 to AAF58593, and AAF21589 to  
 CC AAF21606 represent PCR primers which are used in the exemplification of  
 CC the present invention. The NMB genome and fragments from it have  
 CC antibacterial activity, and can be used in vaccines and gene therapy.  
 CC *Neisseria* nucleic acids, proteins and/or antibodies which binds to the  
 CC proteins can be used in compositions for treating or preventing infection  
 CC due to *Neisseria* bacteria or as a diagnostic reagent for detecting the  
 CC presence of *Neisseria* bacteria or of antibodies raised to *Neisseria*  
 CC bacteria. Computers, computer memory, computer storage medium or computer  
 CC databases can be used in a search to identify open reading frames (ORFs)  
 CC or coding sequences within the NMB genome. The DNA sequences provide  
 CC further opportunities to find antigenic or immunogenic proteins which are  
 CC more effective in vaccines than the outer membrane proteins currently  
 CC used.

XX Sequence 349980 BP; 84410 A; 84863 C; 94187 G; 86520 T; 0 other;

Alignment Scores:

Pred. No.: 2.41e+05 Length: 349980  
 Score: 46.00 Matches: 11  
 Percent Similarity: 65.38% Conservatives: 6  
 Best Local Similarity: 42.31% Mismatches: 9  
 Query Match: 35.66% Indels: 0  
 DB: 21 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x AAF21607 (1-349980)

QY 3 AlaleuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaVal 22

DB 111891 GCTTTGAAGCGCGTTTGGAAACTCGAAGCGCGCAAAATCCGAGCGGCAGCATG 111950

QY 23 LysAlaAspAsnAlaAla 28

DB 111951 GAAGCGGCTGAAGCGGCT 111968

RESULT 32

AAH68525  
 ID AAH68525 standard; DNA: 349980 BP.

XX AAH68525;

XX 26-SEP-2001 (first entry)

XX C glutamicum coding sequence fragment SEQ ID NO: 7060.

XX *Corynebacterium*; amino acid synthesis; vitamin; saccharide;

XX organic acid synthesis; ds.

XX *Corynebacterium glutamicum*.

XX EP1108790-A2.

XX 20-JUN-2001.

XX 18-DEC-2000; 2000EP-0127688.

XX 16-DEC-1999; 99JP-0377484.

QY 24 AlaAspAsnAla 27  
DB 576 TTCGAGACGCT 565  
RESULT 29  
ID AAV74485/c  
XX AAV74485 standard; DNA; 4549 BP.  
XX AC AAV74485;  
XX DT 16-MAR-1999 (first entry)  
XX DE Staphylococcus aureus contig SEQ ID #174.  
XX KW Computer readable medium; vaccine; S.aureus infection; immunodetection;  
KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;  
KW skin infection; surgical wound infection; scalded skin syndrome;  
KW toxic shock syndrome; ds.  
XX STaphylococcus aureus.  
XX FH Key Location/Qualifiers  
FT misc\_feature 121..180  
FT /tag= a  
FT /note= "these bases represent a line of missing text in  
FT the sequence listing in the specification. They  
FT are included to maintain the nucleotide numbering  
FT given in the specification for this DNA sequence".  
FT misc\_feature 1921..1980  
FT /tag= b  
FT /note= "these bases represent a line of missing text in  
FT the sequence listing in the specification. They  
FT are included to maintain the nucleotide numbering  
FT given in the specification for this DNA sequence".  
FT misc\_feature 3721..3780  
FT /tag= c  
FT /note= "these bases represent a line of missing text in  
FT the sequence listing in the specification. They  
FT are included to maintain the nucleotide numbering  
FT given in the specification for this DNA sequence".  
XX EP786519-A2.  
XX 30-JUL-1997.  
XX 07-JAN-1997; 97EP-0100117.  
XX 05-JAN-1996; 96US-0009861.  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA;  
XX Rosen CA;  
XX WPI; 1997-374922/35.  
XX Polynucleotide(s) and proteins derived from Staphylococcus aureus  
XX stored on computer readable medium and used in the production of  
XX anti-S.aureus vaccines  
XX Claim 1; Page 849-852; 3271pp; English.  
XX This sequence represents one of 5191 Staphylococcus aureus DNA sequences  
XX of the invention. The DNA sequences are recorded on a computer readable  
XX medium, preferably selected from a floppy or hard disk, random access  
XX memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using  
XX the S.aureus DNA sequences allows putative functions to be assigned so  
XX that protein-encoding or regulatory regions of commercial, therapeutic or  
XX industrial importance can be obtained. Specifically, sequences which are  
XX likely to encode antigens have been identified and these polypeptides can  
XX be used in a vaccine composition against S.aureus infection. The  
XX polypeptides can also be used in a kit for the immunodetection of

CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,  
CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,  
CC skin and surgical wound infections, scalded skin syndrome, toxic shock  
CC syndrome, etc. Organisms transformed with the DNA sequences can be used  
CC for recombinant production of the polypeptides. The new DNA sequences  
CC (and their fragments) are useful as primers or probes for isolating  
CC homologues of any of the S.aureus DNA sequences contained on the  
CC computer readable medium.  
XX SQ Sequence 4549 BP; 1387 A; 858 C; 698 G; 1421 T; 185 other;  
Alignment Scores:  
Pred. NO.: 1.03e+03 Length: 4549  
Score: 46.00 Matches: 12  
Percent Similarity: 60.71% Conservative: 5  
Best Local Similarity: 42.86% Mismatches: 7  
Query Match: 35.66% Indels: 4  
DB: 18 Gaps: 1  
US-09-847-539A-6\_COPY\_59\_86 (1-28) x AAV74485 (1-4549)  
QY 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluAlaAla 20  
DB 2765 TCAGATATATTGGAGCGGT-----GACATCTTGCACAAATGAAAAAGTAGGCC 2718  
QY 21 ValValIysAlaAspAsnAlaAla 28  
DB 2717 ATCGTGATATATATATATGTTGCA 2694  
RESULT 30  
AA81479/c  
ID AAA81479 standard; DNA; 69936 BP.  
XX AC AAA81479;  
XX DT 04-DEC-2000 (first entry)  
XX DE N. meningitidis partial DNA sequence gnm\_27 SEQ ID NO:27.  
XX KW Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;  
KW antigen; vaccine; diagnosis; infection; antibacterial; identification;  
KW Meningococcus B; MenB; ds.  
XX OS Neisseria meningitidis.  
XX PN WO200022430-A2.  
XX PD 20-APR-2000.  
XX PF 08-OCT-1999; 99WO-US23573.  
XX PR 09-OCT-1998; 98US-0103794.  
XX PR 30-APR-1999; 99US-0132068.  
XX PA (CHIR ) CHIRON CORP.  
XX PI Frazer CM, Hickey B, Peterson J, Tettelin H, Venter JC;  
PI Masignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;  
PI Rappuoli R, Pizza M;  
XX WPI; 2000-318079/27.  
XX Isolated nucleotide sequences of Neisseria meningitidis which can be  
XX used in the diagnosis and treatment of N. meningitidis infection and  
XX other Neisserial infections, for example, N.gonorrhoea -  
XX Claim 7; Page 547-567; 1760pp; English.  
XX The present invention describes methods of obtaining immunogenic  
XX proteins from Neisseria genomic sequences. AAA81453 to AAA8414  
XX represent specifically claimed Neisseria meningitidis genomic DNA  
XX sequences; AAA81260 to AAA81303 and AAB25620 to AAB25663 represent  
XX Neisseria DNA sequences and their corresponding proteins; AAA81254 to







/clone\_lib="C. reinhardtii CC-1690, normalized, Lambda Zap II"  
 /note="vector: pBluescript II SK-; Site\_1: EcoRI; Site\_2: XhoI; This library, constructed by John Davies and Jeffrey McDermott, combines cDNAs from CC-1690 cells grown to mid-log phase in TAP (acetate-containing) medium in the light, TAP medium in the dark, HS (minimal) medium in ambient levels of CO<sub>2</sub> and HS medium bubbled with 5% CO<sub>2</sub>. PolyA mRNA was purified from each sample, pooled and cDNA synthesized. The cDNA was directionally cloned into Lambda Zap II (Stratagene) in the EcoRI (5') and XhoI (3') sites. pBluescript II SK- plasmids were excised from the lambda Zap clones by superinfection with Exassist (Stratagene) phage. The library was normalized using method 4 described in Bonaldo et al (1996) Genome Research 6: 791-806."

BASE COUNT 153 a 161 c 206 g 130 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 166 Length: 650  
 Score: 51.00 Matches: 11  
 Percent Similarity: 62.96% Conservative: 6  
 Best Local Similarity: 40.74% Mismatches: 10  
 Query Match: 39.53% Indels: 0  
 DB: 10 Gaps: 0

US-09-847-539a-6\_COPY\_59\_86 (1-28) x BE337641 (1-650)

Qy 2 AspalaleuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaVal 21  
 ::: :::: ||| ||::||::||| ||::||::||| ||::||::||| ||::||::|||

Db 507 GAGAGATCGAGCTGCTGGTGAAGACCGACGACCTGCGCAACCGAGCGGAGCAGTT 566

Qy 22 VallysAlaAspAsnAla 28

||| ||::|||

Db 567 CCARAAGAGCGGCGGAGCT 587

RESULT 34

BI633216/c

LOCUS

DEFINITION SD27244.Sprime SD Drosophila melanogaster Schneider L2 cell culture  
 pOT2 Drosophila melanogaster cDNA clone SD27244 5 similar to  
 CGT289: Fban0007289 GO:[ ] located on: 2L 22B8-22B8:: 05/23/2001,  
 mRNA sequence.

ACCESSION

BI633216

VERSION

BI633216.1 GI:15535426

KEYWORDS

EST.

SOURCE

ORGANISM

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 660)

Harvey, D., Brokstein, P., Hong, L., Evans-Holm, M., Su, C., Tsang, G.,

Lewis, S. and Rubin, G.M.

BDGP/HMI Drosophila EST Project

Unpublished (2001)

Contact: Stapleton, M.

BDGP

Lawrence Berkeley National Lab

One Cyclotron Rd, Berkeley, CA 94720, USA

Fax: 510 486 6798

Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu

hit genomic AE003584: arm:2L (1824960,2149443)

estimated-cyto:22B4-22D2: 05/23/2001

Plate: SD.272 row: D column: 8

High quality sequence stop: 529.

Location/Qualifiers

1..660

/organism="Drosophila melanogaster"

/db\_xref="taxon:7227"

/clone="SD27244"

/clone\_lib="SD Drosophila melanogaster Schneider L2 cell

culture pOT2"

/lab\_host="DH5-alpha"  
 /note="Vector: pOT2; Site\_1: EcoRI; Site\_2: XhoI; Sized  
 fractionated cDNAs were directly ligated into pOT2.  
 Plasmid cDNA library."

BASE COUNT 148 a 189 c 158 g 163 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 170 Length: 660  
 Score: 51.00 Matches: 11  
 Percent Similarity: 53.85% Conservative: 3  
 Best Local Similarity: 42.31% Mismatches: 12  
 Query Match: 39.53% Indels: 0  
 DB: 10 Gaps: 0

US-09-847-539a-6\_COPY\_59\_86 (1-28) x BI633216 (1-660)

Qy 2 AspalaleuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaVal 21  
 ||| :::: ||| ||::||::||| ||::||::||| ||::||::||| ||::||::|||

Db 226 GATCGCGGAATCATATAGAACACACACAGCAGCAGGAGGAGGAACAAGCGCG 167

Qy 22 VallysAlaAspAsnAla 27

|||::: ||| |||

Db 166 GTACGAGGAGATCGTCC 149

RESULT 35

AI062753/c

LOCUS

DEFINITION GH02073.Sprime GH Drosophila melanogaster head pOT2 Drosophila  
 melanogaster cDNA clone GH02073 Sprime, mRNA sequence.

ACCESSION

AI062753

VERSION

AI062753.1 GI:3338592

KEYWORDS

EST.

SOURCE

fruit fly.

ORGANISM

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 666)

Harvey, D., Brokstein, P., Hong, L., Evans-Holm, M., Su, C., Tsang, G.,

Lewis, S. and Rubin, G.M.

BDGP/HMI Drosophila EST Project

Unpublished (2001)

Contact: Stapleton, M.

BDGP

Lawrence Berkeley National Lab

One Cyclotron Rd, Berkeley, CA 94720, USA

Fax: 510 486 6798

Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu

hit genomic sequence DS06378

Plate: 20 row: G column: 1

High quality sequence stop: 566.

Location/Qualifiers

1..666

/organism="Drosophila melanogaster"

/db\_xref="taxon:7227"

/clone="GH02073"

/clone\_lib="GH Drosophila melanogaster head pOT2"

/sex="male and female"

/dev\_stage="adult"

/lab\_host="DH5 - alpha"

/note="Organ: head; Vector: pOT2; Site\_1: EcoRI; Site\_2:

XhoI; Sized fractionated cDNAs were directly ligated into

pOT2. Plasmid cDNA library."

BASE COUNT 153 a 192 c 157 g 164 t

ORIGIN

Alignment Scores:

Pred. No.: 172 Length: 666

Score: 51.00 Matches: 11

Percent Similarity: 53.85% Conservative: 3

Best Local Similarity: 42.31% Mismatches: 12

Query Match: 39.53% Indels: 0

```

VERSION      AI062828.1  GI:3338667
KEYWORDS     EST.
SOURCE       fruit fly.
ORGANISM     Drosophila melanogaster
             Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
             Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
             Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE    1 (bases 1 to 633)
AUTHORS      Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
             Lewis,S. and Rubin,G.M.
TITLE        BDGP/HMI Drosophila EST Project
JOURNAL      Unpublished (2001)
COMMENT      Contact: Stapleton, M.
             BDGP
             Lawrence Berkeley National Lab
             One Cyclotron Rd, Berkeley, CA 94720, USA
             Fax: 510 486 6798
             Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
             hit genomic sequence DS06378
             Plate: 21 row: G column: 1
             High quality sequence stop: 567.

FEATURES     source
             1..633
             /organism="Drosophila melanogaster"
             /db_xref="taxon:7227"
             /clone="CH02173"
             /clone_lib="GH Drosophila melanogaster head pot2"
             /sex="male and female"
             /dev_stage="adult"
             /lab_host="DH5 - alpha"
             /note="Organ: head; Vector: pot2; Site_1: EcoRI; Site_2:
             XhoI; Sized fractionated cDNAs were directly ligated into
             pot2. Plasmid cDNA library."
BASE COUNT   140 a 182 c 153 g 158 t
ORIGIN
Alignment Scores:
Pred. No.:    160          Length:    633
Score:        51.00       Matches:    11
Percent Similarity: 53.8%   Conservative: 3
Best Local Similarity: 42.31% Mismatches: 12
Query Match:   39.53%     Indels:    0
DB:            9          Gaps:      0

US-09-847-539a-6_COPY_59_86 (1-28) x AI062828 (1-633)
QY  2  AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaVal 21
|||  ::: ||| ||||| ::: ||||| |||
Db  227 GATCGCGGAATCATATAGAACACACACAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 168
QY  22  ValLysAlaAspAsnAla 27
|||::: ||| |||
Db  167 GTACGAGGAGATCGTGCC 150

RESULT 32
AA391497/c 643 bp mRNA linear EST 19-APR-2001
LOCUS      LD10172.5prime LD Drosophila melanogaster embryo Bluescript
DEFINITION Drosophila melanogaster cDNA clone LD10172 5prime, mRNA sequence.
ACCESSION  AA391497
VERSION     AA391497.1  GI:2044472
KEYWORDS    EST.
SOURCE      fruit fly.
ORGANISM    Drosophila melanogaster
             Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
             Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
             Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE    1 (bases 1 to 643)
AUTHORS      Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
             Lewis,S. and Rubin,G.M.
TITLE        BDGP/HMI Drosophila EST Project
JOURNAL      Unpublished (2001)
COMMENT      Contact: Stapleton, M.

```

```

BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
Plate: 101 row: F column: 12
High quality sequence stop: 463.

FEATURES     source
             1..643
             /organism="Drosophila melanogaster"
             /db_xref="BDGP_EST:BDcIn009443"
             /db_xref="taxon:7227"
             /clone="LD10172"
             /clone_lib="LD Drosophila melanogaster embryo Bluescript"
             /sex="male and female"
             /dev_stage="0 to 24 hours mixed stage embryonic"
             /lab_host="SOLR"
             /note="Organ: embryo; Vector: Bluescript SK; Site_1: EcoRI
             ; Site_2: XhoI; Constructed using Stratagene ZAP-CDNA
             Synthesis kit. Oligo dT-primed and directionally cloned at
             EcoRI and XhoI in Bluescript SK(+/-)"
BASE COUNT   140 a 185 c 154 g 164 t
ORIGIN
Alignment Scores:
Pred. No.:    163          Length:    643
Score:        51.00       Matches:    11
Percent Similarity: 53.85%   Conservative: 3
Best Local Similarity: 42.31% Mismatches: 12
Query Match:   39.53%     Indels:    0
DB:            9          Gaps:      0

US-09-847-539a-6_COPY_59_86 (1-28) x AA391497 (1-643)
QY  2  AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaVal 21
|||  ::: ||| ||||| ::: ||||| |||
Db  213 GATCGCGGAATCATATAGAACACACACAGAGGAGGAGGAGGAGGAGGAGGAGG 154
QY  22  ValLysAlaAspAsnAla 27
|||::: ||| |||
Db  153 GTACGAGGAGATCGTGCC 136

RESULT 33
BE337641 650 bp mRNA linear EST 14-JUL-2000
LOCUS      894047D11.y1 C. reinhardtii CC-1690, normalized, Lambda Zap II
DEFINITION Chlamydomonas reinhardtii cDNA, mRNA sequence.
ACCESSION  BE337641
VERSION     BE337641.1  GI:9210726
KEYWORDS    EST.
SOURCE      Chlamydomonas reinhardtii.
ORGANISM    Chlamydomonas reinhardtii
             Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
             Chlamydomonadaceae; Chlamydomonas.
REFERENCE    1 (bases 1 to 650)
AUTHORS      Grossman,A., Davies,J., Federapfel,N., Harris,E., Lefebvre,P.,
             McDermott,J.P., Silflow,C., Stern,D. and Surzycki,R.
TITLE        Analyses of the Chlamydomonas reinhardtii Genome: A Model,
             Unicellular System for Analyzing Gene Function and Regulation in
             Vascular Plants; project phase 2
JOURNAL      Unpublished (2000)
COMMENT      Contact: Elizabeth H. Harris
             DCMB Box 91000
             Duke University
             Durham, NC 27708-1000, USA
             Tel: 919 613 8164
             Fax: 919 613 8177
             Email: chlamy@duke.edu.
             Location/Qualifiers
             1..650
             /organism="Chlamydomonas reinhardtii"
             /strain="CC-1690 wild type mt+ 21gr"
             /db_xref="taxon:3055"

```



```

/clone_lib="RPCI-23"
/sex="Female"
/lab_host="DH10B"
/Note="Organ: Kidney/Brain; Vector: pBac3 6; Site 1:
ECORI; Site 2: EORI; Female C57BL/6J mouse kidney and/or
brain genomic DNA was isolated and partially digested
with a combination of EcoRI and EcoRI Methylase. Size
selected DNA was cloned into the pBac3.6 vector at the
ECORI sites. The ligation products were transformed into
DH10B electrocompetent cells (BRL Life Technologies)."

BASE COUNT      172 a  155 c  148 g  140 t
ORIGIN

Alignment Scores:
Pred. No.:      154      Length:      615
Score:          51.00    Matches:      11
Percent Similarity: 68.18%  Conservative: 4
Best Local Similarity: 50.00%  Mismatches: 7
Query Match:      39.53%    Indels:      0
DB:              12        Gaps:        0

US-09-847-539a-6_COPY_59_86 (1-28) x AZ698116 (1-615)

Qy      1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaLa 20
|||||  |||||  |||  ::|||  |||||  ::|||  ::
Db      522 TCAGATCCTCTGGAGAGACTTAAGAATCAGACAGATTCCTTAAAGGCGACAGAGAGAG 463

Qy      21 ValVal 22
|||||

Db      462 GTAGTG 457

RESULT 30
BB201301/c
LOCUS      BB201301      622 bp      mRNA      linear      EST 19-OCT-2001
DEFINITION BB201301 RIKEN full-length enriched, 0 day neonate thymus Mus
           musculus cDNA clone A430036B19 3', mRNA sequence.
ACCESSION BB201301
VERSION    BB201301.2 GI:16271557
KEYWORDS
SOURCE     house mouse.
           Mus musculus.
REFERENCE  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
           1 (bases 1 to 622)
AUTHORS    Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.,
           Hiramoto,K., Horii,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
           M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,, Sasaki
           Okazaki,Y., Okido,T., Saiko,R., Sakai,C., Sakai,K., Sano,H.,, Sasaki
           D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
           Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
           Muramatsu,M. and Hayashizaki,Y.
           RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
TITLE      Unpublished (2001)
JOURNAL
COMMENT    On Jun 30, 2000 this sequence version replaced gi:8866254.
           Contact: Yoshihide Hayashizaki
           Laboratory for Genome Exploration Research Group, RIKEN Genomic
           Sciences Center(GSC), Yokohama Institute
           The Institute of Physical and Chemical Research (RIKEN)
           1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
           Tel: 81-45-503-9222
           Fax: 81-45-503-9216
           Email: genome-res@gsc.riken.go.jp,
           URL:http://genome.gsc.riken.go.jp/
           Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
           M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
           Normalization and subtraction of cap-trapper-selected cDNAs to
           prepare full-length cDNA libraries for rapid discovery of new
           genes. Genome Res. 10 (10), 1617-1630 (2000)
           wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
           Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura
           S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
           Hayashizaki,Y.
           RIKEN integrated sequence analysis (RISA) system--384-format

```

```

sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Alzawa
K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
Hayashizaki,Y.
Computational Analysis of Full-Length Mouse cDNAs Compared with
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp/) for
further details.
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
FEATURES             Location/Qualifiers
     source           1..622
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone_lib="RIKEN full-length enriched, 0 day neonate
                     thymus"
                     /tissue_type="thymus"
                     /dev_stage="0 day neonate"
                     /lab_host="DH10B"
                     /note="Site 1: SalI; Site 2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAGGATCCAGAGCTCTTTTCTTTTCTTTTCTTTTCTTTT
                     3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of normalization
                     to Rot -10.0 and subtraction to Rot -459.0. Second
                     strand cDNA was prepared with the primer adaptor of
                     sequence [5' GAGAGAGAGATCTCGAGTAAATTAATTAATTCCTCCCTCC
                     3']. cDNA was cleaved with XhoI and BamHI. Vector: a
                     modified pBluescript KS(+) after bulk excision from Lambda
                     FLC I."
BASE COUNT      177 a  163 c  139 g  143 t
ORIGIN

Alignment Scores:
Pred. No.:      156      Length:      622
Score:          51.00    Matches:      11
Percent Similarity: 68.18%  Conservative: 4
Best Local Similarity: 50.00%  Mismatches: 7
Query Match:      39.53%    Indels:      0
DB:              9        Gaps:        0

US-09-847-539a-6_COPY_59_86 (1-28) x BB201301 (1-622)

Qy      1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaLa 20
|||||  |||||  |||  ::|||  |||||  ::|||  ::
Db      502 TCAGATCCTCTGGAGAGACTTAAGAATCAGACAGATTCCTTAAAGGCGACAGAGAGAG 443

Qy      21 ValVal 22
|||||

Db      442 GTAGTG 437

RESULT 31
AI062828/c
LOCUS      AI062828      633 bp      mRNA      linear      EST 19-APR-2001
DEFINITION GH02173:5prime GH Drosophila melanogaster head POT2 Drosophila
           melanogaster cDNA clone GH02173 5prime, mRNA sequence.
ACCESSION    AI062828

```







```
/db_xref="taxon:7227"
```

```

Qy 2 AspaLaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaVal 21
   |||
   ::: ||| ||||| ||| ::: ||||| |||
Db 211 GATCGGCGAATCATATGAGAACACACAGACGACGAGGAGGAGGAGGAGGAGGCG 152

Qy 22 ValGluAlaAspAsnAla 27
   |||::: ||| |||
Db 151 GTACGAGGAGGATCGTGCC 134

RESULT 18
AW659168/c 548 bp mRNA linear EST 25-APR-2001
LOCUS 96120 MARC 1BOV Bos taurus cDNA 5', mRNA sequence.
DEFINITION AW659168
VERSION AW659168.1 GI:7424995
KEYWORDS EST.
SOURCE COW.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 548)
AUTHORS Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett
,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
Perteau,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
Keele,J.W.
TITLE Sequence evaluation of four pooled-tissue normalized bovine cDNA
libraries and construction of a gene index for cattle
JOURNAL Genome Res. 11 (4), 626-630 (2001)
MEDLINE 21180013
COMMENT Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred
v0.980904 e. Vector identified by cross_match with the 'minscore 18
and -minmatch 12 options.
PCR Primers
FORWARD: AGGAAACAGCTATGACCAT
BACKWARD: GTTTCGCCAGTCACGACG
Plate: 83 row: M column: 13
Seq primer: ATTTAGGTGACACTATAG.
FEATURES
source
Location/Qualifiers
1..548
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC 1BOV"
/tissue_type="pooled"
/lab_host="DH10B"
/Note="Vector: pcMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from lymph node, ovary,
fat, hypothalamus, and pituitary."
BASE COUNT 132 a 143 c 151 g 122 t
ORIGIN
|||
Pred. No.: 131 Length: 548
Score: 51.00 Matches: 11
Percent Similarity: 60.87% Conservative: 3
Best Local Similarity: 47.83% Mismatches: 9
Query Match: 39.53% Indels: 0
DB: 9 Gaps: 0

US-09-847-539a-6_COPY_59_86 (1-28) x AW659168 (1-548)

Qy 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVallys 23
   |||
   ||| ||| |||::: ||| ||||| ||| |||:::
Db 538 CTTTCTCTCCTTGTCTCCAGGCTATGCCGTCAGGAGGAGGAGGAGGACCGTCGCG 479

Qy 24 AlAspAsn 26
   |||:::

```

```

Db 478 GAGGATGAC 470

RESULT 19
BM431282/c 553 bp mRNA linear EST 31-JAN-2002
LOCUS 1Ducl12E08 Bos taurus Duodenum #1 library Bos taurus cDNA, mRNA
DEFINITION BM431282
ACCESSION BM431282.1 GI:18453004
VERSION BM431282
KEYWORDS EST.
SOURCE COW.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 553)
AUTHORS Hansen,C., Fu,A., Meng,Y., Li,C., Okine,E., Senses,C.W., Gordon
,P.M.K. and Moore,S.S.
TITLE Gene Expression Profiling of the Bovine Gastrointestinal Tract
JOURNAL Unpublished (2002)
COMMENT Contact: Dr. Stephen Moore
Beef Genomics Laboratory
Dept of AFNS, University of Alberta
410 Agri/For, Dept of AFNS, U of A, Edmonton, AB, T6G 2P5, Canada
Tel: 780 492 0169
Fax: 780 492 4265
Email: smoores@afns.ualberta.ca
Insert Length: 553 Std Error: 0.00
POLYA=No.
FEATURES
source
Location/Qualifiers
1..553
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="Bos taurus Duodenum #1 library"
/tissue_type="Smooth muscle"
/cell_type="Simple columnar epithelial"
/dev_stage="Young adult"
/lab_host="XLI-BlueMRF'strain"
/Note="Organ: Intestine/duodenum; Vector: Uni-22APXR;
Site_1: EcoRI; Site_2: Xho I"
BASE COUNT 158 a 124 c 139 g 132 t
ORIGIN
|||
Pred. No.: 132 Length: 553
Score: 51.00 Matches: 11
Percent Similarity: 60.87% Conservative: 3
Best Local Similarity: 47.83% Mismatches: 9
Query Match: 39.53% Indels: 0
DB: 10 Gaps: 0

US-09-847-539a-6_COPY_59_86 (1-28) x BM431282 (1-553)

Qy 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVallys 23
   |||
   ||| ||| |||::: ||| ||||| ||| |||:::
Db 142 CTTTCTCTCCTTGTCTCCAGGCTATGCCGTCAGGAGGAGGAGGAGGACCGTCGCG 83

Qy 24 AlAspAsn 26
   |||:::
Db 82 GAGGATGAC 74

RESULT 20
AU206045 558 bp mRNA linear EST 17-JUL-2001
LOCUS AU206045 unpublished oligo-capped cDNA library, stage L1
DEFINITION AU206045
ACCESSION AU206045.1 GI:14838902
VERSION AU206045
KEYWORDS EST.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidea
; Rhabditidae; Peloderinae; Caenorhabditis.

```















US-09-847-539A-6_COPY_59_86 (1-28) x BM396025 (1-861)	
QY	2 AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluAlaAlaVal 21       ::                  ::         ::
Db	501 GATGCTCAACAAGCTCTCTAAGACAACACTGATCTCTCAATGCTGCCGAAGAAGCTCTT 560
QY	22 ValLysAlaAsp 25 
Db	561 GCCAATGCTGAA 572
RESULT 3	
LOCUS	BJ012116
DEFINITION	BJ012116 MF01SSA cDNA Oryzias latipes cDNA clone MF01SSA171D09 5', mRNA linear EST 05-DEC-2001
ACCESSION	BJ012116
VERSION	BJ012116
KEYWORDS	EST.
SOURCE	BJ012116.1 GI:17354928
ORGANISM	Japanese medaka.
REFERENCE	Oryzias latipes
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha; Belontiiformes; Adrianichthyidae; Oryziinae; Oryzias.
TITLE	1 (bases 1 to 613)
COMMENT	Kohara,Y., Shin-i,T., Kimura,T., Narita,T., Jindo,T. and Takeda,H. Medaka EST Project in Takeda's lab Unpublished (2001) Contact: Tadasu Shin-i National Institute of Genetics 1111 Yata, Mishima, Shizuoka 411-8540, Japan Tel: 81-559-81-6856 Fax: 81-559-81-6855 Email: tshini@genes.nig.ac.jp.
FEATURES	Location/Qualifiers
SOURCE	1..613
	/organism="Oryzias latipes"
	/strain="Hd-r"
	/db_xref="taxon:8090"
	/clone="MF01SSA171D09"
	/clone_lib="MF01SSA cDNA"
	/sex="mixture of female and male"
	/tissue_type="whole embryo"
	/dev_stage="segmentation stage 20 - 25"
BASE COUNT	134 a 136 c 239 g 104 t
ORIGIN	
Alignment Scores:	
Pred. No.:	72.9 Length: 613
Score:	53.00 Matches: 12
Percent Similarity:	60.71% Conservative: 5
Best Local Similarity:	42.86% Mismatches: 11
Query Match:	41.09% Indels: 0
DB:	10 Gaps: 0
US-09-847-539A-6_COPY_59_86 (1-28) x BJ012116 (1-613)	
QY	1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluAlaAla 20       ::    ::          ::
Db	8 TCAGAGCGCGGCGACGAGCGCGCGACGAGCGGGCGCTTTAACCGAGGAGGCGCGC 67
QY	21 ValVallysAlaAspAsnAlaA 28 ::: ::
Db	68 GCTTTAACCGAGGAGGCGGCT 91
RESULT 4	
LOCUS	BJ011078
DEFINITION	BJ011078 MF01SSA cDNA Oryzias latipes cDNA clone MF01SSA153F11 5', mRNA linear EST 05-DEC-2001
ACCESSION	BJ011078
ORIGIN	

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 13, 2002, 02:12:56 : Search time 332.257 Seconds  
(without alignments)  
1137.418 Million cell updates/sec

Title: US-09-847-539a-6\_COPY\_59\_86  
Perfect score: 129  
Sequence: 1 SDALEALADQTDALQSEAAVVKADNAA 28

Scoring table: BLOSUM62  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters: -DEV-xmlh  
-O/cqn2\_1/USPTO\_spool/US09847539/runat\_10102002\_092548\_3435/app\_query.fasta\_1.526  
-DB=EST -QFMT=FASTAP -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=40 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09847539@cqn\_1\_1\_763@runat\_10102002\_092548\_3435 -NCPU=6 -ICPU=3  
-NO\_XLPHY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THRADS=1 -XGAPOP=60 -XGAPEXT=60 -Fgapop=6 -Fgapext=7  
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estopl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_htc:\*  
12: gb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_inv:\*  
15: em\_gss\_pln:\*  
16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	54	41.9	788	10 BM400696	
2	54	41.9	861	10 BM396025	

3	53	41.1	613	10	BJ012116
4	53	41.1	615	10	BJ011078
5	53	41.1	624	10	BJ013747
6	53	41.1	1101	12	CNS000FX5
7	52	40.3	438	9	AA941632
8	52	40.3	553	9	BE056346
9	52	40.3	1163	12	CNS06NI7
10	51	39.5	243	10	BF496493
11	51	39.5	333	9	AA140863
12	51	39.5	383	9	AI552971
13	51	39.5	387	9	AI467576
14	51	39.5	454	9	AA423358
15	51	39.5	491	9	AI405338
16	51	39.5	529	9	AU101404
17	51	39.5	535	9	AA540164
18	51	39.5	548	9	AW659168
19	51	39.5	553	10	BM431282
20	51	39.5	558	9	AU206045
21	51	39.5	563	10	BF486089
22	51	39.5	567	9	AI063297
23	51	39.5	567	9	AI388871
24	51	39.5	567	10	BI637415
25	51	39.5	568	9	AI109571
26	51	39.5	569	9	AA392809
27	51	39.5	581	9	AI388559
28	51	39.5	583	9	AI108834
29	51	39.5	615	12	AZ698116
30	51	39.5	622	9	BB201301
31	51	39.5	633	9	AI062828
32	51	39.5	643	9	AA391497
33	51	39.5	650	10	BE337641
34	51	39.5	660	10	BI633216
35	51	39.5	666	9	AI062753
36	51	39.5	666	12	AZ699109
37	51	39.5	710	10	BI637204
38	51	39.5	797	10	BF494445
39	50	38.8	445	10	BG840448
40	50	38.8	483	10	BF487636
41	50	38.8	495	9	AU208069
42	50	38.8	528	9	AV393274
43	50	38.8	547	10	BG051545
44	50	38.8	550	9	AI919727
45	50	38.8	552	9	AU205634

ALIGNMENTS

RESULT 1  
BM400696

LOCUS  
DEFINITION

5009-0-77-F09.t.1 Chilcoat/Turkewitz cdNA, mRNA sequence.  
Tetrahymena thermophila cdNA, mRNA sequence.

ACCESSION  
VERSION

BM400696  
1 GI:18200749

KEYWORDS  
SOURCE

EST.  
Tetrahymena thermophila.  
Tetrahymena thermophila

ORGANISM

Eukaryota; Alveolata; Ciliophora; Oligohymenophorea;  
Hymenostomatida; Tetrahymenina; Tetrahymena.

REFERENCE

1 (bases 1 to 788)  
Turkewitz, A.P., Karrer, K.M., Jahn, C., Orlas, E., Kirk, K.E., Frankel, J., and Klobutcher, L.

AUTHORS

EST from Tetrahymena thermophila, strain CU428-1, growing cells  
Unpublished (2002)

TITLE  
JOURNAL

Contact: Turkewitz AP  
Molecular Genetics and Cell Biology  
University of Chicago

COMMENT

920 E. 58th Street, Chicago, IL 60637, USA  
Tel: 773 702 4374  
Fax: 773 702 3172  
Email: apturkew@midway.uchicago.edu  
Seq primer: T3.

FEATURES

Location/Qualifiers

```
terminator 10601..10643
/gene="gyra"
gene 10729..12243
/gene="lin0008"
CDS 10729..12243
/gene="lin0008"
/note="similar to cardiolipin synthase"
/codon_start=1
/transl_table=11
/protein_id="CAC95241.1"
/db_xref="GI:16412429"
/translation="MRKLQFLFIADVFLVEVVLINQAAILFLITSGIIQLCGVILIA
IRLLFQORNTSSKAVIAVIFILPVLGTISLVFGRNPATRRPSTAQVMEKAKLINA
IHAIPNTNEKLPRLSLIAHLTSIEPIKGNKIEILTNGEETFPVLLDALRKAENHIH
IQYIYFTDAISTEIRDLVEKAKSGVEVRFMEDGLGSKLSKFLAPLKEAGVSIHA
FDPIASPIVTRTANLRHKIVIDGIGFTGGLNIGEEVRSNTPDFRVWRDTHMKIT
GOAVTELQESFLNWIWENQAGAADOQISEAGLQOQVESPVDVGDEWAQVITYGGPYDK
EKWVRDSMLDIDSAGESVWIVSPYFVPDDEALAVIRRVAMSGVDVVRVPIPGKDRGI
SFHGSNAVYKTIWICAGAKMTAYADDSFVHAKMLVDGTRAAIGTANFEDVRSFRUNHEL
MVFYDESEAMHLKDFKDFKDFSRFLTKDMENKPLLTRIKEVLSLLSPIL"
12247..12777
/gene="lin0009"
RBS 12247..12252
/gene="lin0009"
CDS 12259..12777
/gene="lin0009"
/note="similar to spermidine N1-acetyltransferase"
/codon_start=1
/transl_table=11
/protein_id="CAC95242.1"
/db_xref="GI:16412430"
/translation="MSGDLKRLPLEREDLKFVHRLNNDAKIMSYWFEEPEYAEFVELQE
LYDKHIHQDSERFILDGMVGLVELMEIDYIHRRAEQI IIDPKFQGHGYAVSAT
KLAMKYAFHVLNLHKLYLVVDKVKNEKAIHVVEKVGFIREGELIDFEFFVDGTYHDAIRM
```

## Alignment Scores:

Pred. No.:	6.37e+03	Length:	231450
Score:	51.00	Matches:	10
Percent Similarity:	61.54%	Conservative:	6
Best Local Similarity:	38.46%	Mismatches:	10
Query Match:	39.53%	Indels:	0
DB:	1	Gaps:	0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x AL596163 (1-231450)

```
Qy 1 SerAspAlaLeuAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAla 20
DB 43340 GCAGAACACAAAGACGATAGCTGCGGACCGATATAGATAAGAGATACAGGT 43399
Qy 21 ValValLysAlaAspAsn 26
DB 43400 GGAGTAGAAAAAGAAAC 43417
```

Search completed: October 13, 2002, 03:04:30  
Job time : 503.262 secs

Tierrez,A., Vazquez-Boland,J.A., Voss,H., Wehland,J. and Cossart,P.  
Comparative genomics of *Listeria* species  
Science 294 (5543), 849-852 (2001)  
21537279  
2 (bases 1 to 231450)  
Glaser,P., Frangeul,L. and Rusniok,C.  
Direct Submission  
Submitted (09-JUL-2001) Glaser P., Institut Pasteur, Genomique des  
Microorganismes Pathogenes, 25 rue du Docteur Roux, 75724 Paris  
Cedex 15, FRANCE  
E-mail: pglaser@pasteur.fr  
Phone: +33 (0)1 45 68 89 96, Fax: +33 (0)1 45 68 87 86.  
Location/Qualifiers  
1. -231450  
/organism="Listeria innocua"  
/strain="Clp11262"  
/db\_xref="taxon:1642"  
306. .311  
/gene="dnaA"  
306. .1674  
/gene="dnaA"  
319. .1674  
/gene="dnaA"  
/codon\_start=1  
/transl\_table=11  
/product="Chromosomal replication initiation protein DnaA"  
/protein\_id="CAC95234.1"  
/db\_xref="GI:16412422"

COMMENT  
E-mail: pglaser@pasteur.fr

Phone: +33 (0)1 45 68 89 96, Fax: +33 (0)1 45 68 87 86.

Location/Qualifiers

Source

RBS

gene

CDS

gene

CDS

terminator

gene

CDS

gene

CDS

gene

CDS

gene

CDS

RBS

gene

CDS

/gene="lin0004"  
4645. .4863  
/gene="lin0004"  
/note="similar to B. subtilis YaaA protein"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAC95237.1"  
/db\_xref="GI:16412425"  
/translation="MAETVKINSEFVTLQQLQMLQIDVYSTGGMAKAYLSENTIYINGE  
QNRKGLKLRNGDVLIVPCGVKVKIEQG"

4870. .5982  
/gene="RecF"  
4870. .5982  
/gene="RecF"  
/codon\_start=1  
/transl\_table=11  
/product="RecF protein"  
/protein\_id="CAC95238.1"  
/db\_xref="GI:16412426"  
/translation="MHLESIVLRNERNYENLELSPSVNVFLGSENAQKTNLLEAVL  
MLALAKSHRTTNDKDFIMWEKEERMEGRVVRKQTVPLLELAITQGRKRVNHLLEQK  
KLSQVGNLVNVIAPEDLSLVKGAPVRRRLNMEITGOMQPIYLHNSEYQRILQQR  
NOYLKMLQMKRVDPMLLDILTQEFADVAINLTKRADP IQLEAYAAP IHHQISRLG  
ETLKIYKASVTLNGDDPEVKADLLQKMEISKQREIDRGVTLIGPHRDSLFYINGO  
NYDFGSQGOORTTALSVKLAIEDLIHEETGEVPVLLDDVLSELDDYRQSHLLGAIE  
GKVFQVTTTSTSGIDHNTLRQATTFYVEKGVKKS"

6031. .7971  
/gene="gyrB"  
6031. .7971  
/gene="gyrB"  
/codon\_start=1  
/transl\_table=11  
/product="DNA gyrase subunit B"  
/protein\_id="CAC95239.1"  
/db\_xref="GI:16412427"

/translation="MSEENITNVHESASDYNEDQIVLQGLEAVRRKRRPGMYIGSTSOR  
GLHLVWEIVDNA IDEALAGFCTEITIEADNSITVRDNGRGIPTGINEKIGRPTVE  
VIPTVLHAGGKFGGGYKVSGLHGVSVAALSTSLVYVHRGOKYORFERGDV  
VMDMEVQGETDRTGTVIHFPTPOPIFTETTFDPTLRTRELAFNLNGLTISIEDK  
REEHKVDFHYEGGIRVSEHLNKAQDVHEPPYILEGERDDIMVEISMQNTQFSS  
NISFANNIHTYEGGTHESGFKTALTINDYARRNKLPKDSDNLSGDDVREGLTAI  
ISIKHPDQFQGTNKLCSSEARSITDKLFSALNKMFMENPDVAKKIVEKGVVASR  
ARLAARKAREVARKSSGLEISLPGKLADCSSRNPEISELYIVEGDSAGSKAQGRDR  
LFOATILPRTGKILNVEKARLDRIANEERTIFTAMGTGFGGDFQVSKRYKHLIMT  
DADVDGAHIRTULLTLTRYMRPLDAGTIIYAQPLIOIKRKGKQIEYVYSDGQLEDY  
LASLGDRTKYSIQRYKGLGEMNPEQLMDTMTMPEHRTLLQVNIKDAIDADEFTEMLMG  
DVEPRRFIEDNAYVKNLDV"

8049. .8054  
/gene="gyrA"  
8049. .10643  
/gene="gyrA"  
8066. .10594  
/gene="gyrA"  
/codon\_start=1  
/transl\_table=11  
/product="DNA gyrase subunit A"  
/protein\_id="CAC95240.1"  
/db\_xref="GI:16412428"

/translation="MAETPNQRITELNLMKERTSFLDYAMSVIVARALPDVDRGLKP  
VHRRILYAMNDLGMTSDKAYKKARSIVGEVIGKYHPHGDTAVYFTWVRMAQDFSYRNM  
LVDHGNFGSVGDMAAMRYTEARMSKISMELLRDINKDTIDYADNDYDGSREPVIL  
PARFPNLLVNGSSGIAVGMATNIPTHHLGEVLDGVLALSHDPEISTRLMEHILPGPDF  
PIATMGRSGIRRAYESGRGSITVRGRVDIEEKNKRETIIVITELPQVQNKARLVNR  
IAELAREKIDGITSLNDSRGMRIVIEVRDJSASVIVNNLMKMTALQTPFGINM  
LALVNHPPKVLNKLLEIHYLEHKVIRRRTEFELRKAERALEGLRIALDNIDA  
ITKLKRGKTSDAKEGLMTQFNLSQQAQALDMLRQLTGLEREKIEEYQNILVAL  
INDLKAILLADDERLLEIIRRELEEIKVKYAKRRTEILAGLDVLEDEDLIPEEVAI  
TITKGYIKRLPLSTYRSQRCRGEGICQMSTHEDDFEHLVATSTHDTLLRPTTKGVK  
YRSKGYEPEYCGRTAKGIPIDINLLGIESQBQVNAVNLSEFTDDSYLFTTTHGVKVR  
YTLSQFAKIRQSLRAVELRENDLISVQMTDGSKNMI IATKRGQSITFPENIRVMT  
PTAAGVRGIRLREDDVEYGMVEDEKDLVVTYKGYQKQTPASQYPLRNRRGMGVGT  
VITTEKGNLVAKMTVTGEEIDLMTVSGVLIRFETDVTVSQTRGSAMGVKKLIRLDEDE  
KVATVAKVPKEDEVELEEEIDETLITQVPDESFEADAPGSDIEE"



```

/db_xref="GI:17428525"
/translation="MEAVLIVDSRTGTARQVAQRLAAQSGWALAEVGLRPRAGFSG
DARCVLENLNRASYRYEGRALDGFQEVITAPVWAGRLASPMRIFLADRLPFPVRL
PRCA"
1570..1788
/gene="RSC1510"
/note="RS03794"
1570..1788
/gene="RSC1510"
/function="miscellaneous; unknown"
/note="Product confidence : hypothetical
Gene name confidence : hypothetical
Predicted by Codon_usage
Predicted by Framed"
/codon_start=1
/transl_table=11
/evidence="not_experimental
/product="HYPOTHETICAL PROTEIN"
/protein_id="CAD15212.1"
/db_xref="GI:17428526"
/translation="MCVMAARGAFNAIEETAGLTSTLLAPVLALLQDVASGLSQDEI
SGPIHOVRTASDRSRRRAANLSPSEA"
1839..2207
/gene="RSC1511"
/note="RS03793"
1839..2207
/gene="RSC1511"
/function="miscellaneous; hypothetical/global homology"
/note="Product confidence : hypothetical
Gene name confidence : hypothetical
Predicted by Codon_usage
Predicted by Homology
Predicted by Framed"
/codon_start=1
/transl_table=11
/evidence="not_experimental
/product="CONSERVED HYPOTHETICAL PROTEIN"
/protein_id="CAD15213.1"
/db_xref="GI:17428527"
/translation="MAQLTAAQHQHIALLDQEARLHROLTELESVPADARPAVEP
YEVDLADLAEASERADMENHYTELAQIIVARERLADRGYGLCVDCGEAIFFLRLQ
AQTAQCCVACQKRERRNA"
complement(2254..2532)
/gene="RSC1512"
/note="RS03792"
complement(2254..2532)
/gene="RSC1512"
/function="miscellaneous; unknown"
/note="Product confidence : hypothetical
Gene name confidence : hypothetical
Predicted by Codon_usage
Predicted by Framed"
/codon_start=1
/transl_table=11
/evidence="not_experimental
/product="HYPOTHETICAL PROTEIN"
/protein_id="CAD15214.1"
/db_xref="GI:17428528"
/translation="MKSIEPPGPRPOYSATNLTLTCVANVDGSPACYAVTAEALEDHF
GAKSYIQEDLLLELAEYHRDAEDMARALFKLDHNLVLRSGHFRFGM"
complement(2866..3321)
/gene="rnhA"
complement(2866..3321)
/gene="rnhA"
/EC_number="3.1.26.4"
/function="macromolecule metabolism; macromolecule
degradation; degradation of rna"
/note="Product confidence : probable
Gene name confidence : probable
Predicted by Codon_usage
Predicted by Homology
Predicted by Framed"

```

```

/codon_start=1
/transl_table=11
/evidence="not_experimental
/product="PROBABLE RIBONUCLEASE HI PROTEIN"
/protein_id="CAD15215.1"
/db_xref="GI:17428529"
/translation="MOEVTIVYSDGACKGNPGI GGWGTVLVSGGHEKELFGGBAVTTNN
RMELMAVIEAFRALRPRCVKYVTDQYQKGI SEWLAKGKRWKTADKKPKVKNDDL
WRTLDLVVYTHEVSWHWKGHAGHPGNERADALANKVEIARQAIQAGA"
complement(3365..4138)
/gene="RSC1514"
/note="RS03790"
complement(3365..4138)
/gene="RSC1514"
/function="miscellaneous; hypothetical/partial homology"
/note="Product confidence : hypothetical
Gene name confidence : hypothetical
Predicted by Codon_usage
Predicted by Homology
Predicted by Framed"
/codon_start=1
/transl_table=11
/evidence="not_experimental
/product="HYPOTHETICAL PROTEIN"
/protein_id="CAD15216.1"
/db_xref="GI:17428530"
/translation="MFGFHALQHGPGALRENRIPLIARVIDDIDPGAPIDTLDPAA
DDPGCEADAEPVQORATPRVICRYDELPPASOSIDLVALPHVLFTDDPHEVLRE
VARVLNPEGRVLTGPNPLSLMGMRGRLGTESLPASQMIATRLKWLKLGFG
DIVRPFQGCPPNRTDKWLQRTAFMEKAGRWMPIFGAVYMLQAVKRVSRVLVGPA
WKRKSPTLVPAGTPVATPFGSHSTTKTPPPVAARDGKSGA"
4227..5081
/gene="gloB"
/note="RSC1515; RS03789"
4227..5081
/gene="gloB"

```

Alignment Scores:

Pred. No.:	5..54e+03	Length:	202050
Score:	51.00	Matches:	11
Percent Similarity:	68.42%	Conservative:	2
Best Local Similarity:	57.89%	Mismatches:	6
Query Match:	39.53%	Indels:	0
DB:	1	Gaps:	0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x AL646065 (1-202050)

Qy 9 AspGlnThrAspAlaLeuGlnSerGluGluAlaValValLysAlaAspAsuAla 27  
||||| |||||||::: ||| ||||| |||||:::|

Db 3887 GACCAGGTCGATGCTCGAGGCGAACGCGAGCTCGTAGCGGAGATCACGCG 3943

RESULT 40

AL596163 231450 bp DNA linear BCT 04-DEC-2001

LOCUS Listeria innocua Clip11262 complete genome, segment 1/12.

DEFINITION Listeria innocua Clip11262

ACCESSION AL596163

VERSION AL596163.1 GI:16412421

KEYWORDS Listeria innocua.

SOURCE Listeria innocua

ORGANISM Listeria innocua

REFERENCE 1 (sites) Bacillus/Staphylococcus group; Listeria.

AUTHORS Glaser, P., Frangeul, L., Buchrieser, C., Rusniok, C., Amend, A., Baquero, F., Berche, P., Bloeker, H., Brandt, P., Chakraborty, T., Charbit, A., Chetouani, F., Couve, E., de Daruvar, A., Dehoux, P., Dommann, E., Dominguez-Bernal, G., Duchaud, E., Durant, L., Dussurget, O., Entian, K. D., Fsihi, H., Portillo, F. G., Garrido, P., Gautier, L., Goebel, W., Gomez-Lopez, N., Hain, T., Hauf, J., Jackson, D., Jones, L. M., Kaerst, U., Kreft, J., Kuhn, M., Kunst, F., Kurapkut, G., Madueno, E., Maitournam, A., Vicente, J. M., Ng, E., Nedjari, H., Nordstiek, G., Novella, S., de Pablos, B., Perez-Diaz, J. C., Purcell, R., Rammel, B., Rose, M., Schlueter, T., Simoes, N.,

PARPNNLVNGSSGAVGMATNIPHTHGLGEVIGDVLALSHDPEITIRDMLEYIRPDF  
PTAGMIRSGIRIRAYSGSGITVRGRVDIEKKNGKETIVITEIPQVANKARLVER  
IAELAREKKIDITSLNDESORSGRMIVIEVRDISASVIVNLFKMTALQTTGGINM  
LALYDNPKVLNLKLEILLYLKHQKVIRRTERELPKAEARAHLEGLRALDNIDA  
IILKIRKSTSDVAKEGMLTQFNLSKQQAAILDMRLQRLTGLEREKEEYQNLVAL  
INDUKLADDERLEIRELEEKIKYKADKRREILAGDLVLEDEDLPEEENVAI  
TLTKRGYIKRLPLSTYRSQRRGGIOGMSHEDDFVHLVATSHDTLLFFTTGKV  
YRSQYEVPEYGRKAGPIINLLGISQEQVNAVINSFTDDSYLFTTKHGVVKR  
TTLQFARIGRGLRAVELRENDELISVQMTDGSKNMIIATKHGSIVFPENIRVMG  
RTAGAGCIRLREDEVGMEVLEDEKVLVYVTEKGYGKQTPASQYPLNRGGMGVKT  
VTTEKGNLVMATVTCGEEDMLMTVSGVLIREFEIDTVSQTGRSAMGVKLIRLDEDE  
KVATVAVPKPEDEVELEEEIDETLITQVPDESAPGSDIE"

terminator

10600. 10653

/gene="gyrA"

10712. 12242

/gene="lmo0008"

10712. 10717

/gene="lmo0008"

10728. 12242

/gene="lmo0008"

/note="similar to cardiolipin synthase"

/codon\_start=1

/transl\_table=11

/protein\_id="CAC98223.1"

/db\_xref="GI:16409367"

/translation="MRKLIQFLAVLVFLVEYVLIQAAILFLVTSGIOLCGVIA  
IRLLFPORNTSSKVAWVAVIFILPVLTISLVFGNRPATRKSTAGVKKTKLINA  
IHAIPNNTNEKPLNLSKIAHLSTIEPIKNGKIELTNGESTFPVLLDALKRAENHII  
IOYIFTDATSTIREIDILVEKAGSVEVRFMDGLSSKLGKAPLAPLKAGVSIHA  
FDPIASPIWATNLRNHRKIIVIDQGTGTLNIGEEYRSNTPDFRWRDTHIKIT  
QOAVTELQESPLNDWYFMENAGADGFISSGSKQFSPVDMGDWAQVYVGGPYDK  
EKWVRDSMLDLSAKESVWTVSPVFPDESILAVIRRVANSVDVVRVILPKGDGRI  
SFGNSNAYVTKMIAGAKMYAYADDSFVHAKAMLVDTGTRAAICTANFDVRSFRLNHEL  
MVFLYDESEAHHLKDFKDFEDSRFLTKMDKENKPLLTIRIKEVLSLLSPIL"

## Alignment Scores:

Pred. No.:	5,48e+03	Length:	200050
Score:	51.00	Matches:	10
Percent Similarity:	65.38%	Conservative:	7
Best Local Similarity:	38.46%	Mismatches:	9
Query Match:	39.53%	Indels:	0
DB:	1	Gaps:	0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x AL591973 (1-200050)

QY 1 SeraspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaLa 20

: : : : : ||| ||| : : : : : ||| ||| ||| ||| : : : : : |||

Db 51231 GCTGAACAAAAGAACAACTAAGTGATGAACAGATCGACTTGACCGAGAGATGCAGGT 51290

QY 21 ValValLysAlaAspAsn 26

: : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||

Db 51291 GGAGTTGAAAGAAAGAAC 51308

## RESULT 39

AL646065

LOCUS

DEFINITION Ralstonia solanacearum GM11000 chromosome, complete sequence;

segment 9/19.

ACCESSION AL646065

VERSION AL646065.1

KEYWORDS GI:17428522

SOURCE Ralstonia solanacearum.

ORGANISM Ralstonia solanacearum

Bacteria; Proteobacteria; beta subdivision; Ralstonia group;

Ralstonia.

REFERENCE 1 (bases 1 to 202050)

AUTHORS

Salanoubat,M., Genin,S., Artiguenave,F., Gouzy,J., Mangenot,S.,  
Arlat,M., Billault,A., Bottier,P., Camus,J.C., Cattolico,L.,  
Chandler,M., Choisme,N., Claudel-Renard,C., Cunnac,S., Demange,N.,  
Gaspin,C., Lavié,M., Moisan,A., Robert,C., Saurin,W., Schiex,T.,  
Sigulier,P., Thebaud,P., Whalen,M., Wincker,P., Levy,M.,  
Weissenbach,J. and Boucher,C.A.

Genome sequence of the plant pathogen Ralstonia solanacearum

TITLE

## JOURNAL

REFERENCE 2 (bases 1 to 202050)

AUTHORS Boucher,C.A.

TITLE Direct Submission

## JOURNAL

Submitted (05-DEC-2001) Genoscope and CNRS UMR-8030, 2 rue Gaston  
Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biologie  
Moléculaire des Interactions Plantes-Microorganismes INRA-CNRS,  
BP27, 31326 Castanet-Tolosan Cedex, France, Fondation Jean  
Dausset-CRPH, 27 rue Juliette Dodu, 75010 Paris, France, LMG CNRS  
118 Route de Narbonne, F 31062 Toulouse Cedex 4, Genoscope and INRA  
URGV, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France,  
Laboratoire de Biométrie et Intelligence Artificielle INRA, BP27,  
F31326 Castanet-Tolosan Cedex. Laboratoire de Genétique Cellulaire  
INRA, BP27, F31326 Castanet-Tolosan Cedex  
Christian.Boucher@toulouse.inra.fr  
http://sequence.toulouse.inra.fr/R.solanacearum.html.

## COMMENT

## FEATURES

source

1..202050

Location/Qualifiers

/organism="Ralstonia solanacearum"

/strain="GM11000"

/db\_xref="taxon:305"

complement(210..686)

/gene="Rsc1507"

/note="RS03796"

complement(210..686)

/gene="Rsc1507"

/function="miscellaneous: unknown"

/note="Product confidence : hypothetical

Gene name confidence : hypothetical

Predicted by Codon\_usage

predicted by FrameD"

/codon\_start=1

/transl\_table=11

/evidence=not\_experimental

/product="HYPOTHETICAL PROTEIN"

/protein\_id="CAD15209.1"

/db\_xref="GI:17428523"

/translation="NEPRIDIPLEIFRTOLOALHTLTLOESQERMLKTLKSVEDA  
EORSLEACASAEAKTWPEMALPSLIQCRTEQNMFTWGLMNSQHALQOEAH  
TCADAWRTIASNMGSDPLMAPVRSFLDPLSETRPEPPAPAKKSSATASAT"

complement(863..1255)

/gene="Rsc1508"

/note="RS06024"

complement(863..1255)

/gene="Rsc1508"

/function="miscellaneous: unknown"

/note="Product confidence : hypothetical

Gene name confidence : hypothetical

Predicted by FrameD"

/codon\_start=1

/transl\_table=11

/evidence=not\_experimental

/product="HYPOTHETICAL PROTEIN"

/protein\_id="CAD15210.1"

/db\_xref="GI:17428524"

/translation="MPRAKLOVPAPPDNDPNANHRPAPALASTTETTHAASVATR  
DAFGSRQLASVTPSTGTGTGARGIERLGTGRAPSRMRSSANLAHAPARVAPGWPA  
PLDGVRDSILAHMECALTYTISEAARS"

complement(1259..1579)

/gene="Rsc1509"

/note="RS04827"

complement(1259..1579)

/gene="Rsc1509"

/function="miscellaneous: hypothetical/global homology"

/note="Product confidence : hypothetical

Gene name confidence : hypothetical

Predicted by Codon\_usage

predicted by FrameD"

/codon\_start=1

/transl\_table=11

/evidence=not\_experimental

/product="CONSERVED HYPOTHETICAL PROTEIN"

/protein\_id="CAD15211.1"

Tierrez.A., Vazquez-Boland,J.A., Voss.H., Wehland,J. and Cossart,P.  
Comparative genomics of *Listeria species*  
Science 294 (5543), 849-852 (2001)  
21537279  
2 (bases 1 to 200050)  
Glaser,P., Frangeul,L. and Rusniok,C.  
Direct Submission  
Submitted (06-JUN-2001) Glaser P., Institut Pasteur, Genomique des  
Microorganismes Pathogenes, 25 rue du Docteur Roux, 75724 Paris  
Cedex 15, FRANCE  
E-mail: pglaser@pasteur.fr  
Phone: +33 1 45 68 89 96, Fax: +33 (0)1 45 68 87 86.

## COMMENT

## FEATURES

## source

1. .200050  
Location/Qualifiers

/organism="Listeria monocytogenes"

/strain="EGD-e"

/db\_xref="taxon:1639"

305..310

/gene="dnaA"

305..1673

/gene="dnaA"

318..1673

/gene="dnaA"

/codon\_start=1

/transl\_table=1

/product="Chromosomal replication initiation protein DnaA"

/protein\_id="CAC98216.1"

/db\_xref="GI:16409360"

/translation="MQSIEDIQETLQIVKKNKSPSYDTWMTKSTAHSELTGRTIIS

ANNFRVRLHLEKSYQFIANIQLQETIGRLDFVRFDGEEFNEFTVYKPNPALDEBG

IEGRHVLQPRVDFVYIGSGNFRFAAASLAVAAAPAKAYNPLFIYGVGGLGKTHLM

HAQVYQOQHAKQVMYLSSEKFTNEFISSDRNKTEFTFKYRNVVLLDDITDQFL

AKGTEQEEFFNTLYDEQKIIISDRPKETPTDLRLRSFEMGLIDITPPD

LETRAILKAKADGLDIPNEVMYIANQIDSNIRELEGALIRVVAYSLSVKNQITTA

GLAAEAKDIIPSSKQVITISGIOFAVGEYFHVRLDFKAKKRTKSIAFPQJIAWYL

SRELTDASLPKIDEEFGGRDHTVTHAHEKISQLKTDQVLNDLAEIEKNLRKAQNM

F"

1856..3062

/gene="dnaA"

1856..1860

/gene="dnaA"

1867..3012

/gene="dnaA"

/codon\_start=1

/transl\_table=1

/product="DNA polymerase III, beta chain"

/protein\_id="CAC98217.1"

/db\_xref="GI:16409361"

/translation="MAFVIERDLVQAVNEVTRAISSARTIPILTGIIKIVNDEGVTL

TGSDSDISAEFIPLIENDEIVEVEFSGIYLQSKYFGDIVRRLPEENVEIEVTSNY

QTNISGQASFTLNGLDPMYKLPVTDGKTIKIPYLNKIVRQTVFAVAIEVRP

VLTVGNWIIKENKLSAVATDSRLALREIPLDIDEEYNIVIPGKLSLAKLLDDA

SESIEMTLANNOILFKLDLLFYSLRLEGSYDPTSLRIPTDKSELVINSKAFLOAID

RASLLARENWVVKIWLMTLENGOVEVSSNPSPEGVNSENVSQSFTEGEEKISFNCKY

MDALRAFEGDDIQLISFGSGTMRPFLVRPKDAANPNELLQIIPVRY"

3030..3062

/gene="dnaA"

3104..3109

/gene="lmc0003"

3104..4464

/gene="lmc0003"

3121..4464

/gene="lmc0003"

/note="conserved hypothetical protein"

/codon\_start=1

/transl\_table=1

/protein\_id="CAC98218.1"

/db\_xref="GI:16409362"

/translation="MMKDMTGNPTKLIFLAMPMLIGNLFOQFYTMIDAVIVGKFSV

VDAAAGVATNSVNFPMIILIGMSGIVVAAQYFGKDYDRLKDVATATVAVVFS

AITLTAVAGVLAKPILLRTPANILDSTIFLTLFGLTFLPSLYNGMAAILRALGN

SITPLIFLILSLMNLALDFLVVMDGVRGAATVLSQTPAAATVYIAYRHVPF

MRIERAKFLSTPLLAEMVRIGLPSGLQGSFISIGNMALQSILINGFSSVVAAYTAAAS

## RBS

## gene

## CDS

RIDSITYQGIAGFAASSMAQONIGAGKIDRVREGFWSGIKVWTAISIGITILVOLFF  
ARQFILLFVDSSETEVINIGSVYLLISFLVYVVGILFVYVRETLRGTDGMPVPLANGI  
FELVSRLVIGFISLVYIGVGLWATPWAITATILGVNRYKSGAQKKAVIRRK"  
4631..4637  
/gene="lmc0004"  
Join(4631..4637,4644..4865)  
/gene="lmc0004"  
4644..4865  
/gene="lmc0004"  
/note="similar to B. subtilis YaaA protein"  
/codon\_start=1  
/transl\_table=1  
/protein\_id="CAC98219.1"  
/db\_xref="GI:16409363"  
/translation="NAETVKINSEFVTLQQLQIMDVSTGGMKAYLSENTIYINGE  
QNRGKGLRNGDVILVPGVGKIKQKQ"  
4856..4861  
/gene="RecF"  
4856..4861  
/gene="RecF"  
4869..5981  
/gene="RecF"  
4869..5981  
/gene="RecF"  
/codon\_start=1  
/transl\_table=1  
/product="RecF protein"  
/protein\_id="CAC98220.1"  
/db\_xref="GI:16409364"  
/translation="MHLESIVLNRFRNYENLELEFSPSYNVFLGENAQKTNLLEAVL  
MLAAGSHNTNDKDFIMWEKEAKMEGRIAHQGVLPLELAITQKGRKAYNHLQK  
KLSQVYVGNVNIWIFAPEDISLVKAGFIRRELMGEIQMGPYILHNLSEYORILQOR  
NOYLMQKRRVDPILLDLTEQFADVAINTLKRADFQKLEAYAAADTHHQISRLG  
ETLKBYKASITLNGDDPEVWAKDLQKMSIKQRIDRGVTLIGPHRDSLFYINGQ  
NVQDFSGSQGQRTALS IKLAETDLTHEETGEVPLLLDDVLSELDDYRQSHLLGAIE  
GKVPFVFTTSTSGIDHETLKQATTFYVEKGTVKKS"  
6017..7970  
/gene="gyrB"  
6017..6021  
/gene="gyrB"  
6030..7970  
/gene="gyrB"  
/codon\_start=1  
/transl\_table=1  
/product="DNA gyrase subunit B"  
/protein\_id="CAC98221.1"  
/db\_xref="GI:16409365"  
/translation="MSEENITNVQENASDYNEDQIQVLEGLAVRKPGRMYIGSTSOR  
GLHLVWEIVDNAIDALAGFCEIEITIEDNSITVRDNGRGIPTGINKEIGRPTVE  
VIFTVLHAGKFGGGYKVGSLHGVASVYNALSTSLEVYHREGQKYVQRERGDV  
VMDMEQGTDRYGTIVHFTPDQIFTEFEFDTLRTRELALNRLGLTISIEDK  
REBHKVRKDFHVEGGIRSYVEHLNKAQVHEPIYLEGEDDIMVEISNOYNTFSS  
NIISFANNTHVEGGTHESGFTALT RVINDYARNKLFKSDNNLSGEDVRGLTAI  
ISIKHPDQEGQTKLGNSEARSITDKLFSALNKMENPDVAKKIVEKGVASR  
ALAAKRAREVARKSSGLEISLPGKADCSSRNPESLEYIVEGDSAGSQAQGRDR  
LFQAILPRTGKILNVEKADRLILANEINTIFTAMGTGGGDFVSKSRHKLITMT  
DADVDGAHITLTLFIRYMRPLDAGIYIIAQPLPYQIKHGKQLEYVSDGLEDY  
LASLDGHTKYSIQRYKGLGEMNPEQLMDTMTNPEHRTLLQVNIKDAIDADETFEMLMG  
DRVEPRKFTIEDNAQVVKNDV"  
8048..10653  
/gene="gyrA"  
8048..8053  
/gene="gyrA"  
8065..10593  
/gene="gyrA"  
/codon\_start=1  
/transl\_table=1  
/product="DNA gyrase subunit A"  
/protein\_id="CAC98222.1"  
/db\_xref="GI:16409366"  
/translation="MAETPNORITEINLNKEMRTSFLDVAMSIVIVARALPDVRDLKLP  
VHRRLIYAMNDLGTSDKAYKKARSARIVGEVYHGHGDTAVFTVWRALQDQFSYRNM  
LVDHGNGFSGVDGDMAAAMRYTEARMSKYSMELLRDINKDTIDYADNYDGSEREPVIL

## RBS

## gene

## gene

## CDS

## gene

## RBS

## CDS

## gene

## RBS

## CDS

## terminator

## RBS

## gene

## CDS



```
/protein_id="AAK15440.1"
/db_xref="GI:13194222"
/translaton="MAGASAKVAMLLSLVATYGFAGVYVYVNDLWLPKATWYGOPNG
AGPDNDGKGFNTNYYPPMSMTSCGNEPLFDQKGCACGYQIRCTNNPSCSQPRT
VITDNNYGVARYHFDLSTAFAMARPGNLQDHAGIIDIQFRVPVCHRGLYVN
FHYEAGSNPVYLAFLVFAFKDGTGVQLDVNESLPSGCKPTVMTPMRRSNGSIWRLDA
NHRLOGPFSLRWVSESQTVIAHQVTPANWRANTYVGSQVQFR"
repeat_region
complement(11312..17335)
/rpt_family="(CA)n"
join(21858..22143,22248..22357,22530..23373)
mrna
/gene="OSJNB0014111.3"
21858..23373
/gene="OSJNB0014111.3"
/notes="identical to beta expansin EXPB6 GB:AAF72987
GI:8118430 (Oryza sativa) (Proc Natl Acad Sci U S A
94 (12), 6559 6564 (1997)); EST AU062937, C72255, AU161321
from this gene"
CDS
join(21954..22143,22248..22357,22530..23057)
/gene="OSJNB0014111.3"
/codon_start=1
/product="beta-expansin EXPB6"
/protein_id="AAK15442.1"
/db_xref="GI:13194224"
/translation="MAARMGSKVAAILAILSVLVHSGKCHGPNVYVNSDASVYGSW
LPARATWYAGTGGPDGAGCGFRKNVQYPFSSMTSCGNEPIFKDGKGCSCYQIR
CNKDPSCSGNIETVITDMNYVPVARYHFDLSTAFAMARPGNLQDHAGIIDIQF
RRVPCNPGKINPHVEGDSNPVYFAVLVEYEDLDGVQVQVDMLESKSYGGATGVMT
PMRESGSIWRLDSNHRLOAPFSLRISDSKTLVANNVIPANWSPNSRYSIVQFS"
23102..23121
/rpt_family="(CGAT)n"
complement(26118..26245,26118..26381))
/gene="OSJNB0014111.4"
/notes="predicted by fgenesh"
complement(join(26118..26245,26351..26381))
/gene="OSJNB0014111.4"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAK15444.1"
/db_xref="GI:13194226"
/translation="MPEKQPKPAVAAPQRRGGRGCAADGVHREAGDRPKAQQNW
DGWRREGD"
complement(26200..26222)
repeat_region
/rpt_family="GC-rich"
complement(27602..27629)
repeat_region
/rpt_family="GC-rich"
27980..28118
/rpt_family="putative stowaway type MITE element"
complement(27991..28129)
repeat_region
/rpt_family="putative stowaway type MITE element"
complement(join(29835..30595,31066..31120))
mrna
/gene="OSJNB0014111.15"
complement(29835..31120)
/gene="OSJNB0014111.15"
/notes="predicted by genscan"
complement(join(29835..30595,31066..31120))
CDS
/gene="OSJNB0014111.15"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAK15452.1"
/db_xref="GI:13194234"
/translation="MKEKERKADSAVLGFPPAMERERRTPSAATSLHRRLLTLVD
AIGSKAVCGGGERRRIPQHRPSLFRHQACQRPVGARSFPFPFHSQGTIM
RPFPVRIALRLRPGPLPRRAARLPHLPNPVPPPPASKLPIILYFHGGSFYLFST
DNVYFHASCEAAATVPVAILVSLDRLAPEHLPAAYVSTVWLCDAAAGDPWLAASA
REVAGVAGVAGCLASLSLPSYSGRHHSSPPPLPLPAAQAEREKREGDREKW"
complement(29843..29879)
repeat_region
/rpt_family="(GA)n"
30600..30852
repeat_region
/rpt_family="(CGACA)n"
complement(32120..32144)
repeat_region
/rpt_family="AT-rich"
join(<33752..33943,35026..>35142)
mrna
/gene="OSJNB0014111.13"
```

```
Alignment Scores:
Pred. No.: 3.66e+03 Length: 135509
Score: 51.00 Matches: 13
Percent Similarity: 57.69% Conservative: 2
Best Local Similarity: 50.00% Mismatches: 11
Query Match: 39.53% Indels: 0
DB: 8 Gaps: 0

US-09-847-539a-6_COPY_59_86 (1-28) x AC037426 (1-135509)

Qy 3 AlalaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVal 22
Db 67185 GCGCGGAGGCGCGCGCTACGGAGGCGACGCGCGCTACGGAGGCGCGACGACGCG 67126
Qy 23 LysAlaAspAsnAlaAla 28
Db 67125 GCGCGGATACGGCGCGCG 67108

RESULT 36
AC105899 175115 bp DNA linear INV 10-JAN-2002
LOCUS Drosophila melanogaster, chromosome 2L, region 22C-22C, BAC clone
DEFINITION BACR11F10, complete sequence.
ACCESSION AC105899
VERSION AC105899.1 GI:18104818
KEYWORDS HTG.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 175115)
Celniker,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H.,
Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C.,
Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,
Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M.,
Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,
Ferreira,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A.,
Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattel,B., Moshrefi,A.,
McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,
Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B., Scheeler,F.,
Phouanavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
Stapleton,M., Strong,R., Svirskas,R., Tector,C., Williams,S.M.,
Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
Sequencing of Drosophila chromosome 2L, region 22C-22C
Unpublished (1998)
2 (bases 1 to 175115)
Celniker,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H.,
Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C.,
Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,
Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M.,
Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,
Ferreira,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A.,
Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattel,B., Moshrefi,A.,
McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,
Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B., Scheeler,F.,
Phouanavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
Stapleton,M., Strong,R., Svirskas,R., Tector,C., Williams,S.M.,
Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
Direct Submission
Submitted (10-JAN-2002) Berkeley Drosophila Genome Project, MS
64-121, Lawrence Berkeley National Laboratory, One Cyclotron Road,
Berkeley, CA 94720, US
Sequence submitted by:
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720
This sequence was assembled using end sequences from a whole genome
shotgun and from subclones of this BAC and its neighboring clones.
For further information about this sequence, including its location
```

```

Pred. No.:      2.84e+03      Length:      105823
Score:          51.00      Matches:      11
Percent Similarity: 55.8%      Conservative: 3
Best Local Similarity: 42.31%      Mismatches: 12
Query Match:      39.53%      Indels: 0
DB:              2      Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x AC019871 (1-105823)

QY 2 AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaVal 21
   |||  ::  |||  |||||  ::  |||||  |||
Db 20507 GATCGCGGAATCATATAGAACACACAGACGACGAGGAGGAGGAGGAGGAGGCG 20566

QY 22 ValLysAlaAspAsnAla 27
   |||::  |||  |||
Db 20567 GTACGAGGAGATCGTGCC 20584

RESULT 35
AC037426/c
LOCUS
DEFINITION
Oryza sativa chromosome 10 BAC OSJNB001411 genomic sequence,
complete sequence.
ACCESSION
AC037426
VERSION
AC037426.11 GI:13184072
KEYWORDS
HTG.
SOURCE
Oryza sativa.
ORGANISM
Oryza sativa.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
1 (bases 1 to 135509)
Hsiao,J., Yuan,Q., Ouyang,S., Moffat,K.S., Hill,J.N., Burr,P.C.,
VanKien,S.E., Bowman,C.L., Pai,G., Bowman,C.L., Fujii,C.Y.,
Feldblyum,T.V., Bowman,C.L., Craven,B., Utterback,T.R., Khalak,H.,
Fraser,C.M., Quackenbush,J., White,O., Salzberg,S.L. and
Oryza sativa chromosome 10 BAC OSJNB001411 genomic sequence
Unpublished
2 (bases 1 to 135509)
Buell,R.
Direct Submission
Submitted (08-APR-2000) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
3 (bases 1 to 135509)
Buell,R.
Direct Submission
Submitted (05-JAN-2001) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, rbuell@tigr.org
4 (bases 1 to 135509)
Buell,R.
Direct Submission
Submitted (02-MAR-2001) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
5 (bases 1 to 135509)
Buell,R.
Direct Submission
Submitted (03-MAR-2001) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, rbuell@tigr.org
6 (bases 1 to 135509)
Buell,R.
Direct Submission
Submitted (13-MAR-2001) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, rbuell@tigr.org
7 (bases 1 to 135509)
Buell,R.
Direct Submission
Submitted (29-AUG-2001) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, rbuell@tigr.org
On Mar 2, 2001 this sequence version replaced gi:12039293.
Address all correspondence to:rice@tigr.org

BAC clone OSJNB001411 is from Oryza sativa chromosome 10
The orientation of the sequence is from SP6 to T7 end of the BAC

```

```

clone.
Genes were identified by a combination of several methods: Gene
prediction programs including GENSCAN and Genscan+ (Chris Burge,
http://CCR-081.mit.edu/GENSCAN.html), GeneMarkHM (Mark Borodovsky,
http://genemark.biology.gatech.edu/GeneMark/), Fgenesh
(http://www.softberry.com/), and GeneSplicer (Mihaela Pertea and
Steven Salzberg, contact mpertea@tigr.org), searches of the
complete sequence against a peptide database and the plant EST
database at TIGR (http://www.tigr.org/tdb/tgi.shtml). Annotated
genes are named to indicate the level of evidence for their
annotation. Genes with similarity to other proteins are named after
the database hits. Genes without significant peptide similarity but
with EST similarity are named as unknown proteins. Genes without
protein or EST similarity, that are predicted by more than two gene
prediction programs over most of their length are annotated as
hypothetical proteins. Genes encoding tRNAs are predicted by
tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/Eddy/tRNAscan-SE/).
Simple repeats are identified by RepeatMasker (Arian Smit,
http://ftp.genome.washington.edu/RM/RepeatMasker.html).

FEATURES
            source
            1..135509
            /organism="Oryza sativa"
            /cultivar="Nipponbare"
            /sub_species="japonica"
            /db_xref="taxon:4530"
            /chromosome="10"
            /map="E617685"
            /clone="OSJNB001411"
            /complement(479..617)
            /rpt_family="AT_rich"
            join(3823..4103,4207..4316,4453..4637,4983..5310)
            /gene="OSJNB001411.1"
            /note="Identical to beta expansin EXPB3 GB:AAF72984
            GI:8118423 (Oryza sativa) (Proc Natl Acad Sci U S A
            92 (20), 9245-9249 (1995)); EST AA751536, AU030829 from
            this gene"
            join(3920..4103,4207..4316,4453..4637,4983..5310)
            /gene="OSJNB001411.1"
            /codon_start=1
            /product="beta-expansin EXPB3"
            /protein_id="AAK15453.1"
            /db_xref="GI:13194235"
            /translation="MAFSISKKAANAALFSLVTCVAGARGNFSASDFTADPNWEV
            ARATWYGAPTGAGDDGACGFKNTNQYFSSMTSCGNEPIFKDGKGCSCYQTRCV
            NPACSGNPETVITDMNYPVYSKYHFDLSGAFGAMKPGQNDQURHAGIIDIQPKR
            VPCNPGLKVFHVEGSPNPYFAVLVEYEDGDGVQVDLMANSQSWTPMRESMGS
            IWLDSNHLRTAPFSLRITNESGKQLVASQVIPANWAPNAVYRSEFQYSS"
            complement(7169..7197)
            /rpt_family="GC_rich"
            complement(7933..7962)
            /rpt_family="AT_rich"
            complement(9012..9053)
            /rpt_family="AT_rich"
            complement(11040..11098)
            /rpt_family="AT_rich"
            complement(13525..13554)
            /rpt_family="AT_rich"
            complement(14048..14071)
            /rpt_family="AT_rich"
            join(15881..16100,16205..16314,16671..17602)
            /gene="OSJNB001411.2"
            /note="Identical beta expansin EXPB2 GB:AA616710
            GI:2224915 (Oryza sativa) (Proc Natl Acad Sci
            92 (20), 9245-9249 (1995)); EST D24261, BE041015, AU164552
            from this gene"
            join(15950..16100,16205..16314,16671..17195)
            /gene="OSJNB001411.2"
            /codon_start=1
            /product="beta-expansin EXPB2"

```

(657 aa), fasta scores; opt: 479 z-score: 421.7 E():  
3e-16, 28.2% identity in 333 aa overlap. Contains possible  
N-terminal signal sequence"

/codon\_start=1

/transl\_table=11

/product="putative secreted protein"

/protein\_id="CAA20497.1"

/db\_xref="GI:3451438"

/db\_xref="SPTREMBL:O86543"

TRGAEVAVTRADRODLTVRSTALGGRTVEVRLTDPGNPHDRHROHPTLWL  
LHCCGDTVSWTNDTDAETESLRDLVVMPEAGWNGWYSDWNHGGGDPAMETHT  
KEHLHLERDNGAGSRVAGLSMGCGGALLIYAARHPGMPRTAASGSAHPLNDES  
VDRIWGFAGDNDPLRWGDPVAQRGIQWADHPFLAKRLRSIPVYLSGDDGTGPL  
DAPGATGADNDNRQNHAAELKRVGARHVTHFYGPCTHGMWAYRERELHASLPLM  
LGALRVGD"

6183. .6186

RBS

#### Alignment Scores:

Pred. No.:	784	Length:	30561
Score:	51.00	Matches:	12
Percent Similarity:	60.00%	Conservative:	3
Best Local Similarity:	48.00%	Mismatches:	10
Query Match:	39.53%	Indels:	0
DB:	1	Gaps:	0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x SCLF2 (1-30561)

Qy 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAla 20

Db 15463 GCGGACGCGCTGCAGCCCTCCGACGAGATCGCGGCCACCCATTCGTCGGTCCGTCGAC 15404

Qy 21 ValVallysAlaAsp 25

Db 15403 GTGATCGCGCGGAC 15389

#### RESULT 33

AC005749/c

#### LOCUS

DEFINITION Drosophila melanogaster DNA sequence (Pls DS06378 (D278) and DS03995 (D277)), complete sequence.

AC005749 AC004576 AC004575

AC005749.1 GI:3688071

HTG.

Drosophila melanogaster (Subclones in sac from P1 clones DS06378 (D278) and DS03995 (D277)) DNA.

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 96288)

Celniker,S.E., George,R.A., Galle,R.F., Hoskins,R.A.,  
Svirskas,R.R., Harris,N.L., Agbayani,A., Arcaina,T.T., Baxter,E.,  
Blazej,R.G., Chavez,C., Chew,M., Doyle,C.M., Farfan,D.E.,  
Flanagan,J., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,  
Kim,S.H., Lee,B., Lomotan,M.A., Mak,J.J., Mazda,P., Moshrefi,A.R.,  
Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Punch,E.,  
Snir,E., Twomey,B., Wan,K.H., Zhang,R., Zieran,L.L. and Rubin,G.M.  
Sequencing of Drosophila chromosome 2L, region 22B3-22C1  
Unpublished (1997)

2 (bases 1 to 96288)

Celniker,S.E., George,R.A., Galle,R.F., Hoskins,R.A.,  
Svirskas,R.R., Harris,N.L., Agbayani,A., Arcaina,T.T., Baxter,E.,  
Blazej,R.G., Chavez,C., Chew,M., Doyle,C.M., Farfan,D.E.,  
Flanagan,J., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,  
Kim,S.H., Lee,B., Lomotan,M.A., Mak,J.J., Mazda,P., Moshrefi,A.R.,  
Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Punch,E.,  
Snir,E., Twomey,B., Wan,K.H., Zhang,R., Zieran,L.L. and Rubin,G.M.  
Direct Submission  
Submitted (02-OCT-1998) Berkeley Drosophila Genome Project, MS  
64-121, Lawrence Berkeley National Laboratory, One Cyclotron Road,  
Berkeley, CA 94720, US

#### TITLE

#### JOURNAL

#### COMMENT

Sequence submitted by:

Berkeley Drosophila Genome Project  
Lawrence Berkeley National Laboratory, MS 64-121  
Berkeley, CA 94720

For further information about this sequence, including its location  
and relationship to other sequences, please visit our sequence  
archive Web site (<http://fruitfly.berkeley.edu/sequence/>) or send  
email to [bdg@fruitfly.berkeley.edu](mailto:bdg@fruitfly.berkeley.edu).

Library locations: 42-67, 59-42.

#### FEATURES

source

1. .96288

Location/Qualifiers

/organism="Drosophila melanogaster"

/db\_xref="taxon:7227"

/chromosome="2L"

/map="22B3-22C1"

/clone="Pls DS06378 (D278) and DS03995 (D277)"  
/note="These two Pls were sequenced as a project. DS06378  
(D278) extends from p1 end at bp 1 to p1 end at bp 85081.  
DS03995 (D277) is a bridge clone extending from p1 end at  
bp 57,688 to bp 96,288 minimally overlapping its proximal  
neighbor DS00164 (D269)."

BASE COUNT 26399 a 21521 c 21934 g 26434 t

#### ORIGIN

#### Alignment Scores:

Pred. No.:	2.57e+03	Length:	96288
Score:	51.00	Matches:	11
Percent Similarity:	53.85%	Conservative:	3
Best Local Similarity:	42.31%	Mismatches:	12
Query Match:	39.53%	Indels:	0
DB:	3	Gaps:	0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x AC005749 (1-96288)

Qy 2 AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaVal 21

Db 49770 CATCGCGGATCATATAGAGACACACAGCAGCAGAGGACGAGACACAGCGCG 49711

Qy 22 VallysAlaAspAsnAla 27

Db 49710 GTACGAGGATCGTGCC 49693

#### RESULT 34

AC019871

LOCUS

DEFINITION Drosophila melanogaster, \*\*\* SEQUENCING IN PROGRESS \*\*\*, in ordered

AC019871

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

105823 bp DNA linear HTG 03-JAN-2000

fruit fly.

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 105823)

Submitted (30-DEC-1999) Celera Genomics, 45 West Gude Drive,

Rockville, MD, USA

This sequence was identified as CDM-10211226 by the submitter.

For more information on this record e-mail to [fly@celera.com](mailto:fly@celera.com).

\* NOTE: This is a 'working draft' sequence.

\* This sequence will be replaced

\* by the finished sequence as soon as it is available and

\* the accession number will be preserved.

Location/Qualifiers

1. .105823

/organism="Drosophila melanogaster"

/db\_xref="taxon:7227"

BASE COUNT 29823 a 23528 c 23181 g 29291 t

#### ORIGIN

#### Alignment Scores:

The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous. The length in codons is given for each CDS.

Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database. The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon. Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4768-4778(1994)) and the FramePlot program of Bibb et al., Gene 30:157-66(1984) as implemented at <http://www.nih.gov/jp/jun/cgi-bin/frameplot.pl>.

CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, ttg or (att)) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.

IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions. Cosmid 1F2 lies between 8A6 and 4G2 on the AseI-A genomic restriction fragment.

## FEATURES

## source

```
1. .30561
/organism="Streptomyces coelicolor A3(2)"
/strain="A3(2)"
/db_xref="taxon:100226"
/clone="cosmid 1F2"
1. .936
/gene="SC1F2.01"
```

```
1. .102
/gene="SC1F2.01"
```

## misc\_feature

```
/note="SC1F2.01"
```

```
/note="nominal overlap with cosmid 8A6 from 37344 to 37445"
```

## misc\_feature

```
1. .927
```

```
/gene="SC1F2.01"
```

```
/note="Pfam match to entry PF00232 glycosyl_hydrol,
```

```
Glycosyl hydrolases family 1, score 350.20, E-value
```

```
2.2e-101"
```

```
<1. .936
```

```
/gene="SC1F2.01"
```

```
/note="SC1F2.01, probable beta-glucosidase, partial CDS,
```

```
len >311 aa; similar to many e.g. BGLS_AGRSP
```

```
beta-glucosidase (EC 3.2.1.21) (459 aa), fasta scores;
```

```
opt: 784 z-score: 986.7 E(): 0, 42.5% identity in 315 aa
```

```
overlap. Overlaps and extends SC8A6.25. Contains Pfam
```

```
match to entry PF00232 glycosyl_hydrol, Glycosyl
```

```
hydrolases family 1, score 350.20, E-value 2.2e-101"
```

```
/codon_start=1
```

```
/transl_table=11
```

```
/product="putative beta-glucosidase"
```

```
/protein_id="CAA20492.1"
```

```
/db_xref="GI:3451434"
```

```
/db_xref="SPTREMBL:O86539"
```

```
/translation="AEHFAAAYVAERLGRVAVHWATLNEPLCSNIGHLSEKKNAPG
```

```
QTDLRNARASHLLHGLHQAIAAPQAGVIVNLSITHAATDREEDLAARR
```

```
MDGHTNRWLDPVHGFQPDQMRVYGVLPEDPMDRLIGAEGLWLNLYFPQTVA
```

```
ADPTGAPQVTRVTRAGVPTGMDWEIDAGGTEELLRLTEYGPRLKLYVTENGSSFP
```

```
LVLRPDGVDDEPERQETLTHLAACASARKGAPLAGYFAWLLDNFEWAYGYDKRFG
```

```
LVHVDYATQVTRIKGSGHHYADIIGRANGRRKAA"
```

```
1051. .1236
```

```
/note="SC1F2.02, probable beta-glucanase pseudogene
```

```
fragment, len: 61 aa; this appears to be a gene remnant
```

```
similar to part of e.g. TR:O68641 (EMBL:AF052745)
```

```
beta-1,3-glucanase II from Oerskovia xanthinolytica (435
```

```
aa), fasta scores: opt: 100 z-score: 207.7 E(): 0.00025,
```

```
38.3% identity in 60 aa overlap. There is no start codon"
```

```
/codon_start=1
```

```
/pseudo
```

```
/transl_table=11
```

```
/db_xref="PID:e1316894"
```

```
complement(1275. .2177)
```

## CDS

```
/gene="SC1F2.03c"
complement(1275. .2177)
/gene="SC1F2.03c"
/note="SC1F2.03c, unknown, len: 300 aa; similar to
overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein SC1F2.03c"
/protein_id="CAA20494.1"
/db_xref="GI:3451435"
/db_xref="SPTREMBL:O86540"
/translation="MMFRPALTQVSTLEPWAHELRTESGRWTSLSGCSAAHPDRI
DTSRPHARMYDFLGGCKNDYEDQRAAQPKSAPEVREGVLANRHFMRHVRHVVA
EGVQKILDVGTGLTPENVHQAISVAPGTVAVVNDPIVATHSRALADPDPTAVV
LADLRDPRSLIEHPEVRAVDFDRPVALLVAVHVFVADAQAGVATVLRDALPAGS
YVLVSHATGDVHEDRDAAYVKNATASLNLRSHAAVLDFGDFSLVPEGLVRVTDW
RPEEARPPADPIGVIYGVARKDA"
2299. .2304
/note="possible RBS upstream of SC1F3.04"
2308. .3093
/gene="SC1F3.04"
2308. .3093
/gene="SC1F3.04"
/note="SC1F3.04, unknown, len: 261 aa; similar to upstream
gene SC1F2.03c.E(): 0, 48.8% identity in 260 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein SC1F3.04"
/protein_id="CAA20495.1"
/db_xref="GI:3451436"
/db_xref="SPTREMBL:O86541"
/translation="MTAGTPRPDIDTSKPHPARVYDMLGGKNDYVDORVGFTLPAE
ARGNAARNRAFMHRAAAWAKNGVQDFLGTGIPTEPNLHQIVQAISPAARIVYADN
DPTVLRHAEALLTGPEGVTDLFHADVRQPATILERAREVLDFDRPIALSLLALHFL
PDDEDPYRVVSTLVDTLPRGSYLLSHGTADQHPKQETESAYKKGAIPLMRTRSE
VSPFFCGLELVKPLGLVFATVQEEPAVRRSGFFVAGVARIP"
3223. .4881
/gene="SC1F2.05"
3223. .4881
/gene="SC1F2.05"
/note="SC1F2.05, probable secreted protein, len: 552 aa;
contains N-terminal signal sequence"
/codon_start=1
/transl_table=11
/product="putative secreted protein"
/protein_id="CAA20496.1"
/db_xref="GI:3451437"
/db_xref="SPTREMBL:O86542"
/translation="MKLRKRRLCAATGVVLGALTALAAPPAVASPHPGGHPPTARR
ALPLGAPLPERTTTTLQPGVTLTRVGNADDPALHWTVEVSIIPGGTISPDQAPP
TALKDRASADELAAELRRDGFARVEEVTTPATADYRGTLGMRVTRIGTDSQATAA
ERTRLRAGYSGSAVYWDGATDRGPMHVDVLTIDPRTFRGLDASGPDLENRET
TSALAAASAAATAVNAAGFVLDPRAGAPGVGVYDGRLLSEPVAGRPALVVHDDG
RTELARTWQGRITRAASLPNGVNRVPGILRNCGGAGDPTPTSLPHDVTCTNPD
ELVATPDYGGRTPGCEGNEAVLDAHERVVELRSPRGCTIPRGSSVQNTGERVADLT
ATARVGERVRSATLDDDRMRMSTPGDITVNAAGLVDRGVHVTPTADGMVHPGD
PSWYGVWIKHNRPTVAGVDAAGRTVLVTDAGRTSDSLGSLGESAEVARSGLRDAY
NLDGGSTVMAEAGVANSFSDAAGERPVGDALLILPHRRHRTS"
4908. .4991
/note="12 degenerate repeats of AGGCAC(C /A)"
5005. .5008
/note="possible RBS upstream of SC1F2.06"
5014. .6042
/gene="SC1F2.06"
5014. .6042
/gene="SC1F2.06"
/note="SC1F2.06, probable secreted protein, len: 342 aa;
some similarity to secreted antigenic proteins 85A, 85B
and 85C from Mycobacteria e.g. A85B_MYCLE (327 aa), fasta
scores: opt: 325 z-score: 247.2 E(): 1.6e-06, 28.5%
identity in 319 aa overlap and to secreted antigenic
protein PSI from Cornebacterium glutamicum CSp1_CORGL
```

## repeat\_region

## RBS

## gene

## CDS









```
gene
/transl_table=11
/product="Vng6208c"
/protein_id="AAG20870.1"
/db_xref="GI:10584263"
/translation="MKSQLQKGVAVGVAFVLVTVLVGGGLAWNPVQEGNIEVVKEMGA
STGTLEFGANVIVPIKQSTAVVPVPOEYTMANEKQGEAARDSDSEVLVNDGVSVN
VDVIRVVKNTAEATFYDEYKDVSOAEARIKIRPTODVLRTEGGDIDTTEIYTGAGO
KOMAAYKALETAVSGSLIEAVOLRNIKLPGYADAVEKKVEKONIEKKNSIO
VAKAEARKVQAKGEARENEIVAESLKNFELIKIRYIEALNQDNSTIYVGAQSQAG
ITLTKNDVEENKS"
1016..1159
/gene="VNG6209H"
CDS
1016..1159
/gene="VNG6209H"
/feature="hypothetical protein"
/codon_start=1
/transl_table=11
/product="Vng6209h"
/protein_id="AAG20871.1"
/db_xref="GI:10584264"
/translation="MAGRLIVFGLLWLGAIALVTVGAPRYFSERQRLSHERMYDKRH
DDE"
1543..2808
/gene="gabt"
CDS
1543..2808
/gene="VNG6210G"
/feature="gabt"
/feature="gabt"
/codon_start=1
/transl_table=11
/product="gamma-aminobutyrate aminotransferase"
/protein_id="AAG20872.1"
/db_xref="GI:10584265"
/translation="MAVTHAEGIEVTDAGGTYTLDCLAGACTLALGHNHNPVATLELR
LLADEPLHTDITTPKEAFVDFALFSLPASPANTAKVQFCSPAGTDIAEALKLTK
TATNRSVLAFHGGYHGTAGALSMDGIDDAKEPVSGMLPNVHPLPYDPQYRHPFGLP
AGEHSITYTEHAISPERGFTTPAAVVELVOEGGVNPAPDAWQELRRIITREHD
IPLVDEOTQNLGTGETWSEFHAGIEPDVVTCSKAVGGGLPLAVLVNEDGLWEPG
AHACTFRGNLQAMAGRVYIEHVIEQDLPAHADXTGARLSRALTLDLEADFVQVDRG
RGLMGEVFAPDAAGDERPPDGLAAAVKQACFERGLIITETGGRSAVAFVLPPLI
VSSHOIDIAARFSDAVTAVLSSSAEVS"
2919..4379
/gene="bdb"
CDS
2919..4379
/gene="bdb"
/codon_start=1
/transl_table=11
/product="L-2,4-diaminobutyrate decarboxylase"
/protein_id="AAG20873.1"
/db_xref="GI:10584266"
/translation="MEACADAVLDVIGGAHGPYSGASYDRLRLADTTALPAEGHPL
EDVLAAREVDYLANAVHPSDETCAHLQCPMPVPALAAEALITASNQSLDSFDQAPAA
TPVERLIADLTALVELGPAAGVVTGGTETSNHQALLARLDXYETVFGOSVREHGL
PAAQDLRLCSADAHETAQSAALLGLGEDAVVTIPTDGRHRMDATALRETVERLND
AKKIPFVVGATGTTDFGSDIPLSAVDVAATHDLWHDVAAGFAGALAVSDQHSERLS
GIDRADSVAVDFHKLFTFOPIACGALLVADGESFELMSNNAAYLNPAGDAPVNLVAKST
RTTRDALPKYVAFVTRTLGRDGLGALVDRTLALADDVAGGLRADPAFELACEPTLNAY
TFVRVPYREHAHCDPGPWADHVTEAARQLFDAGTAVVARTTVDDRAHVKTTLMNPRT
TVADITQSVLVAFKGAHAATIEAADRTAPITTAAGGQ"
4376..6199
/gene="luca"
CDS
4376..6199
/gene="VNG6212G"
/feature="luca"
/feature="luca"
/codon_start=1
/transl_table=11
/product="iron transport protein A"
/protein_id="AAG20874.1"
/db_xref="GI:10584267"
/translation="MTGVGAVADTPAAHAESATLHAFLNCYLRETDAGEVVSRAVPD
```

```
GVPAQAGRADRRVVRVPPAOSTTTFVAPLRVASATGRHLFAVPAYAVPCVDPDAGCTP
SPIDAAGVASLARRELA LDGDDPTAGRDLLSERVLASQTATFEFVAARDTPRDRHLYD
TTVSFRDAEQSLVYGHMHPTNSRTGIPPHQESRYAPELRAGFOLRVFADPALVET
WAAGDASATEQLSVVVADATTVPPESEATLVPHWQADYLHAQPHVAAAIAARDELRD
LGAFPTPTPTSSVTRLSPDAPFMYKGS LAVEITNAERTNKRRLARGLAVSELLAT
GFGDAFPTFPFAFSTIEDPAATLTLDGCAESGFVAIRENPFAGDDTVNQPVAALC
QDGLDGPAAARIVSRIATREGRTESAWAREFEOYLAHTVRPVVLYDFOVVGLEAH
QOQTLVRLDDCRPTAGFYRONQGYFFPASOAADVDDMLPGVCERADTVCPDAIADER
LRYIVFNNAAGVYNALGAGGVYDERALLAVLRLTELSLAHEPADSTLVLSALLTERR
VPCKANLITRFEERDELEADLETQSYTAIENPLVTRVDTQ"
6196..6906
/gene="lucB"
CDS
6196..6906
/gene="lucB"
/feature="lucB"
/codon_start=1
/transl_table=11
/product="iron transport protein B"
/protein_id="AAG20875.1"
/db_xref="GI:10584268"
/translation="MTAPLPRITSDYAYRRPDPQIDRTVAFRRVALDRDLGRHLHAWLT
SDHVTYWSLGAFLPAFQAIGDHLDEDDHTVAYVGCIDGPMYSYVERVWVAADPLADH
YDADDNQGIIHILIGPPEYLGQCGCTALLRAMTALAFRAHAATDRVVAEPDATNDAAIA
AFERAGFPRETFREPHEDKDATRVVCDRERFVGAFSPCAIAANGPPADGPAVNHDDA
THRCPSGEGHNSLEADR"
6903..8336
/gene="hxyA"
CDS
6903..8336
/gene="hxyA"
/feature="hxyA"
/codon_start=1
/transl_table=11
/product="monooxygenase"
/protein_id="AAG20876.1"
/db_xref="GI:10584269"
/translation="MTESVHDLVGVGLGPNFLAALIEGSDDDVPLDAAFLBQKPEFY
WHDGMLIEGTTLEVPFLADLTADPTNPYSYLNLRENDRLRYEYFFTFQFLPRREY
YRDLVYVGVSGSRPYVPPAFRDLPAFVHTAAYREHRAAVCDADSVAVIGSGSAAE
VWHDLLTRSTSDRSRPVDMVTWSAGFPFMEYSKLGLQHTPEYDYVVDLPQPVKDE
LIPNODILYKGIIDPETSRIYDTLYEQSIGGRDPNVGMIAATEVGAPEPTGGCRYVD
CRHQTDTGFAFETDAVILTCYHRPTPGFLSSLEPRLARODQDRDLAVDREYGVGDL
EGGLFVQNAELTHGVGAPDGLGAYRNAVILDRLLADAPYPVDRDTVFQDPAPAYAA
SHAPGGRVDRDDPTTTPPAAGQPNHHQ"
8362..10188
/gene="lucC"
CDS
8362..10188
/gene="lucC"
/feature="lucC"
/codon_start=1
/transl_table=11
/product="iron transport protein C"
/protein_id="AAG20877.1"
/db_xref="GI:10584270"
/translation="MNTYEDTLQDAVTAATWATYRSLLAKLLREFTYEELIAPERTT
DTRGRYRLVLPAAADADTVRYREAEERLMDTVFPDTIERAPADASWATATDP
TRFLADARETIDATETTTGHLIREYNTLLADAHIEARSAPESFADCSYAESEMTG
HPWLTFRNGRLGMDYEDYDYAPERRQTTTIGWLAVRTDRVTFTVSGITPTGLLRGE
LQDQYDFMATLADKGLDQPEYTLPLPVHNYQWETTITSYASELAADAILVPLNGADS
YLPQSVRTLNFVDAPEKHVKNPKMIRNTLVWRGLPGDRTEAARVTEYITDIDRGG
PFLRDDCELVLPGEGAVNVDHPAFSELEDNAYQYDELLGAVRESDTDAIDPQQAV
TLAALLHEADGTAVISERFSPLESAWLTLEFDTLLPLLHLYRYGTVFSPHGEN
TIVILDEEFVPTRLGKDFVDDVNI SDQPLPELADLADLRVLASEPNGLCQFVFC
GLFVGVRYLSDVLTAHHGDETEFTQVRAITAYQAEFPELSDFETFDLLQPTFT
KLC LNRRLTEGYS DTSRRPAAHEHGTVTNALHEVAGDADE"
complement(10189..10842)
/gene="trkA3"
complement(10189..10842)
/gene="trkA3"
```



```
/protein_id="CAC49491.1"
/db_xref="GI:15140978"
/translation="MCYYPEQPRGKWAEDARRMVELGSLSVIRGEFAWKIEPRSGE
FHWLEDAIDVLGKAGUKYILGTPTAAPKWLNYRPEILPVDATGAVRKGARRHY
CFSSRRSEARITTEAMARYGHIYVHAWQTDNEYGDHDTIYSYSAEAYGAFRLWL
AERYSIDELNRAAGTSFWSNRYOSFEEIDLPNNLVEEPSPTHGVDYFRFSDDQVKS
NKAQVDITRAHSPQRPVTHNPMSQNTDFHYRVGEDIDIASWDVYPMGGLNGLRAAK
DKHYLRVGDPOQAFNHDLPVAVGRGVWMEQPGPVNNAHNSPADGCMVRLWTH
LAYAGVDMYSYFWRQAPFAQEQFHAGLLLPNSEADQGYLEVAEYVAKRLEGEV
RCKAQVALVLYESRKAIRVLPGQKSYSAVAALDMSYTVARLGVDFICQSHDIAG
YKLLADPLVIAEAEFVERTRADAKVVGFGARSGKTRDMHPIPEGLPGPLAKLIDIS
VSRVSEPEFHSETVLYGNEAYEAGGWRTVTSVIAESDFGEYRNGAPALVSDKA
RYLVAANGALLDKVIGDALGWAGLESPLDGLDLRVTTRGNLRFAFNFRIPAEVPAG
ASAEYVGGRTLKPDVAIWE"
/complement(5426..6460)
/gene="SMB20967"
/complement(5426..6460)
/gene="SMB20967"
/function="MISCELLANEOUS; Not classified regulator"
/notes="Product confidence : putative"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative transcriptional regulator, lacI family
protein"
/protein_id="CAC49492.1"
/db_xref="GI:15140979"
/translation="MQRTTIARETGSLSTATSVKALKHSPQVRPETAVALDAAER
VGYELNHVGLRTGKYVAAIMTAPGKENEWEVYLAQLLGSFWALBESYRVS
LYAVDFEESQETISQVSMKADGIIISGRADDPRIRLMQEAEPVYVYGMSTHNA
PHAYVDANQMIKSLVRLERHRIALLNPKPOFAYGVVRLSEYKLEKAGLQPY
DPALVAGRLTPAFGRNVMTMSALADPTAYICANEATALGAFSGPHARGLVHGRDA
VINATDLNVSQYFAPPTISYPLIGPESALLKFLIRMEGEPPDALQTLMLPDLE
RSDRLRPR"
/complement(6468..8180)
/gene="SMB20968"
/complement(6468..8180)
/gene="SMB20968"
/function="MISCELLANEOUS; Hypothetical/Partial homology"
/notes="Product confidence : hypothetical"
Gene name confidence : hypothetical"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="HYPOTHETICAL PROTEIN"
/protein_id="CAC49493.1"
/db_xref="GI:15140980"
/translation="MLFDIDHVPFSRGRFLTSTMRVPGDGERALYLRVSGDGR
PSLGRCLRVFLDGEPPAATRFQSLDQVARTGEGIVREVGDLHRTGNAGV
RFLVGEGRDYVYRTPGGECCLVAAGENVKPIPRGLGELAVSGKWRDRSTVSFTL
SGEFTFGNVDFRALLPPEPGSGFEEAAVATAEFGAWYRSIPAGVPGQEAORLAA
YLLWANTVPAEGLVTRPAIYMSKNGMINISWDNAFSAVGAADEELAFQCPAAIFD
HQDASGLLPDYNDREALFATPKPPVHGVAVSCWARENPAELTPERRAYLRDAIGRQV
SYWTHGRAGNVRLPSYFVGHNDSDNASFFAEGGPLVSPDLVFLILACBALANLE
DDPKAARWLRIADLAQLVTLTGTFETFAARLAADPERILPGESIQFQPLLIGSR
LPAAQCQLVRLADEVGAFITDWPATSPRPFYEDDGYWRGPTWAPTYTLMDGLRR
QGCALAREIAQKFCALANAHGMAENDFARSGLRDRFAFWTSVAVYLLLAQSLRDEK
PVKS"
/complement(8184..8996)
/gene="SMB20969"
/complement(8184..8996)
/gene="SMB20969"
/function="CELL PROCESSES; Transport of small molecules;
carbohydrates, organic acids, alcohols"
/notes="Product confidence : putative"
Gene name confidence : hypothetical"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative sugar uptake ABC transporter permease
protein"
/protein_id="CAC49494.1"
/db_xref="GI:15140981"
```

```
/translation="MRKTVLYATLILLSALFIAPRYFTMTAVKSTAEYLRPPPLWLP
SEWHWENFAAASKQPCFTYLSNSLIVVVLSTIGOLLSSLVAFGFAFRFRPGRDALE
VLLATMIPWDVYKMIPLMYENMLCNWLTLPVLPAYFADAFVFLRLQYIMTIPM

Alignment Scores:
Pred. No.: 5,72e+03 Length: 303100
Score: 52.00 Matches: 11
Percent Similarity: 77.27% Conservative: 6
Best Local Similarity: 50.00% Mismatches: 5
Query Match: 40.31% Indels: 0
DB: 1 Gaps: 0

US-09-847-539a-6_COPY_59_86 (1-28) x RME603646 (1-303100)

QY 4 LeuGUlAlaLeuAlaAspInThrAspAlaLeuGlnSerGluGUlAlaAlaValVallYs 23
DB 38610 ATGGATGATCTGCCGACGGAACGACCGATATTCGTCGTCGACCGCGTGATGCGG 38551
QY 24 AlaAsp 25
DB 38550 CACGAC 38545

RESULT 27
LOCUS AY061146/c 2151 bp mRNA linear INV 08-NOV-2001
AY061146 Drosophila melanogaster LD12115 full length cDNA.
ACCESSION AY061146
VERSION AY061146.1 GI:16768949
KEYWORDS FLI_CDNA.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
Stapleton,M., Brokstein,P., Hong,L., Abgayani,A., Carlson,J.,
Champe,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E., George,R.,
Gonzalez,M., Guarin,H., Li,P., Liao,G., Miranda,A., Mungall,C.J.,
Nunoo,J., Pacleb,J., Paragas,V., Park,S., Phouanavong,S., Wan,K.,
Yu,C., Lewis,S.E., Rubin,G.M. and Celniker,S.
Direct Submission
Submitted (30-OCT-2001) Berkeley Drosophila Genome Project,
Lawrence Berkeley National Laboratory, One Cyclotron Road,
Berkeley, CA 94720, USA
Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory
Berkeley, CA 94720
This clone was sequenced as part of a high-throughput process to
sequence clones from Drosophila Gene Collection 1 (Rubin et al.,
Science 2000). The sequence has been subjected to integrity checks
for sequence accuracy, presence of a polyA tail and contiguity
within 100 kb in the genome. Thus we believe the sequence to
reflect accurately this particular cDNA clone. However, there are
artifacts associated with the generation of cDNA clones that may
have not been detected in our initial analyses such as internal
priming, priming from contaminating genomic DNA, retained introns
due to reverse transcription of unspliced precursor RNAs, and
reverse transcriptase errors that result in single base changes.
For further information about this sequence, including its location
and relationship to other sequences, please visit our web site
(http://fruitfly.berkeley.edu) or send email to
cdna@fruitfly.berkeley.edu.
Location/Qualifiers
1..2151
/organism="Drosophila melanogaster"
/strain="y: cn bw sp"
/db_xref="taxon:7227"
/map="2288-2288"
1..2151
/gene="CG7289"
/notes="alignment with genomic scaffold AE003584"
/db_xref="FLYBASE:FBgn0031379"

FEATURES
source
gene
```







```

QY      24 AlaSpAsn 26      169427 bp      DNA      linear      INV 07-JUN-2001
LOCUS   BACR08101, complete sequence.
DEFINITION Drosophila melanogaster, chromosome 2L, region 28B-28B, BAC clone
AC007416 BACR08101, complete sequence.
AC007416 AC007416.12 GI:14327741
KEYWORDS  HTG.
SOURCE    fruit fly.
ORGANISM  Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 169427)
Celniker,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H.,
Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C.,
Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,
Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M.,
Dodson,K., Dorsett,Y., Doup,L.E., Doyle,C., Garg,N.S., George,R.A.,
Ferreira,S., Frise,E., Galle,R.F., Garg,N.S., Hostin,D., Howland,T.J.,
Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A.,
McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,
Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B., Scheeler,F.,
Phouanavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
Stapleton,M., Strong,R., Svirskas,R., Tector,C., Williams,S.M.,
Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
Sequencing of Drosophila chromosome 2L, region 28B-28B
Unpublished
2 (bases 1 to 169427)
Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.M., Farfan,D.E., Galle,R.A., George,R.A., Harris,N.L.,
Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
Rubin,G.M.
Direct Submission
Submitted (28-APR-1999) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA
On Jun 7, 2001 this sequence version replaced gi:7143379.
Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720
This sequence was assembled using end sequences from a whole genome
shotgun and from subclones of this BAC and its neighboring clones.
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive Web site (http://www.fruitfly.org/sequence/) or send email
to bdg@fruitfly.berkeley.edu.
Location/Qualifiers
1..169427
/organism="Drosophila melanogaster"
/strain="y; cn bw sp"
/db_xref="taxon:7227"
/chromosome="2L"
/map="28B-28B"
/clone="BACR08101 (D605)"
/clone_lib="RPCI-98 (Roswell Park Cancer Institute
Drosophila melanogaster BAC library, partial EcoRI in
pBAC3.6)"
/note="Alternate"
BASE COUNT 51061 a 35853 c 34873 g 47640 t
ORIGIN

```

---

```

Alignment Scores:
Pred. NO.:      3.13e+03      Length:      169427
Score:          52.00         Matches:      11
Percent Similarity: 65.22%      Conservative: 4
Best Local Similarity: 47.83%      Mismatches: 8
Query Match:     40.31%      Indels:      0
DB:              3           Gaps:          0

US-09-847-539a-6_COPY_59_86 (1-28) x AC007416 (1-169427)

QY      4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVallys 23
LOCUS   Drosophila melanogaster genomic scaffold 142000013386055 section 12
DEFINITION Of 63, complete sequence.
AC003619 AC003619.2 GI:10728652
KEYWORDS  HTG.
SOURCE    fruit fly.
ORGANISM  Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,
George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
Sutton,G.G., Wortman,J.R., Yeaman,M.D., Zhang,Q., Chen,L.X.,
Brandon,R.C., Rogers,Y.H., Blazer,R.G., Champe,M., Pfeiffer,B.D.,
Wan,K.H., Doyle,C., Baxter,E.G., Heit,G., Nelson,C.R., Gabor
Miklos,G.L., Abri,J.F., Agbayani,A., An,H.J.,
Andrews-Pfannkuch,C., Baldwin,D., Balley,R.M., Basu,A.,
Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,
Benos,P.V., Berman,B.P., Bhattacharya,D., Bolshakov,S., Borkova,D.,
Botchan,M.R., Bouck,J., Brokstein,P., Brottier,P., Burtis,K.C.,
Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferreira,S.,
Feischmann,W., Foster,C., Gabrielian,A.E., Garg,N.S.,
Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,
Guan,P., Harris,M., Harris,N.L., Harvey,D., Helman,T.J.,
Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,
Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,
Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C.,
McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C.,
Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
Nusskern,D.R., Pacieb,J.M., Palazzolo,M., Pittman,G.S., Pan,S.,
Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,
Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I.,
Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,
Stapleton,M., Strong,R., Sun,E., Svirskas,R., Tector,C., Turner,R.,
Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wasserman,D.A.,
Weinstock,G.M., Weissbach,J., Williams,S.M., Woodage,T.,
Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F.,
Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,
Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,
Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
The genome sequence of Drosophila melanogaster
Science 287 (5461), 2185-2195 (2000)
20196006
TITLE
JOURNAL
MEDLINE

```

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 36993)  
Adams, M. and Venter, J. C.  
Direct Submission  
Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,  
Rockville, MD, USA  
This sequence was identified as CDW:10212995 by the submitter.  
For more information on this record e-mail to fly@celera.com.  
\* NOTE: This is a 'working draft' sequence.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.

FEATURES  
source  
1. .36993  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
BASE COUNT 9781 a 8675 c 8725 g 9812 t  
ORIGIN  
|||||  
Alignment Scores:  
Pred. No.: 649 Length: 36993  
Score: 52.00 Matches: 11  
Percent Similarity: 65.22% Conservative: 4  
Best Local Similarity: 47.83% Mismatches: 8  
Query Match: 40.31% Indels: 0  
DB: 2 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x AC018007 (1-36993)

QY 4 LeuGluaLeuAlaAspGlnThrAspAlaLeuGlnSerGluAlaAlaValLys 23  
|||||  
Db 30206 CTGGAAGCTCTGTTCAAGCAGGTGGAGTCTCTCCAGCGCGAGTTCCTCTGTGGCA 30265  
|||||

QY 24 AlaAspAsn 26  
|||||  
Db 30266 GCCGACAAT 30274  
|||||

RESULT 23  
AC011910  
LOCUS  
DEFINITION  
Drosophila melanogaster clone RPC198-2017, \*\*\* SEQUENCING IN  
PROGRESS \*\*\*, 39 unordered pieces.  
AC011910  
VERSION  
AC011910.1 GI:6056126  
KEYWORDS  
HTG; HTGS\_PHASE1.  
SOURCE  
fruit fly.  
ORGANISM  
Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 133155)

REFERENCE  
AUTHORS  
Muzny, D.M., Adams, C.C., Bailey, M., Barbara, J., Blankenburg, K.,  
Bodota, B., Bock, J., Bowie, S., Brooks, A., Buhay, C., Bunac, C.,  
Burkett, C., Burrows, J., Carter, M., Chacko, J., Chen, Z., Cox, C.,  
David, R., Delgado, O., Deshazo, D., Ding, Y., Domah-Rashid, N.,  
Dugan-Rocha, S., Durbin, K.J., Fernandez, C., Ferraguto, D.,  
Forcum-Tansey, J., Frantz, P., Ganesh, R., Gorrell, J.H., Gorrell, L.L.,  
Guevara, W., Harris, K., Hernandez, J., Hodgson, A., Hogues, M.,  
Holloway, C., Hosak, H., Jackson, L.E., Jackson, L., Jia, Y., Jones, M.,  
Kelly, S., Kondejewski, N., Kong, Y., Kovar, C., Leal, B., Li, Z.,  
Lichtarge, O., Liu, J., Liu, W., Logan, O., Lu, J., Lucier, R.,  
Martin, R., Martinez, C., McLeod, M.P., Mei, G., Morgan, M., Morris, S.,  
Nash, S., Nelson, A., Nguyen, R., Nguyen, N., Nguyen, S., Oswal, G.,  
Parish, B., Paxton, S., Payton, B., Perez, L., Pu, L.L., Quiles, M.,  
Reiter, D., Rives, M., Samuel, S., Say, J., Scherer, S., Shah, B.,  
Shen, H., Simon, M., Sparks, A., Stamps, A., Sucgang, R., Tabor, P.,  
Taylor, T., Vasquez, L., Vinson, R., Vo, O., Wabnah, M., Watlington, S.,  
Weinstock, G., Weinstock, I.R., Williamson, A., Worley, K., Wren, J.,  
Wrensford, G., Yu, W., Zhou, X., Nelson, D. and Gibbs, R.

TITLE  
JOURNAL  
Unpublished  
REFERENCE  
2 (bases 1 to 133155)

AUTHORS  
TITLE  
JOURNAL

Worley, K.C.

Direct Submission

Submitted (16-OCT-1999) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

## COMMENT

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 39 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 854: contig of 854 bp in length  
\* 855 1695: contig of 841 bp in length  
\* 1696 2535: contig of 840 bp in length  
\* 2536 3358: contig of 823 bp in length  
\* 3359 4171: contig of 813 bp in length  
\* 4172 5125: contig of 954 bp in length  
\* 5126 6105: contig of 980 bp in length  
\* 6106 6952: contig of 847 bp in length  
\* 6953 7603: contig of 651 bp in length  
\* 7604 8574: contig of 970 bp in length  
\* 8574 9725: contig of 1151 bp in length  
\* 9725 10972: contig of 1247 bp in length  
\* 10972 12519: contig of 1548 bp in length  
\* 12520 14371: contig of 1852 bp in length  
\* 14372 15264: contig of 893 bp in length  
\* 15265 16905: contig of 1641 bp in length  
\* 16906 19536: contig of 2631 bp in length  
\* 19537 21604: contig of 2068 bp in length  
\* 21605 22985: contig of 1381 bp in length  
\* 22986 24363: contig of 1378 bp in length  
\* 24364 26424: contig of 2061 bp in length  
\* 26425 28205: contig of 1780 bp in length  
\* 28205 31216: contig of 3011 bp in length  
\* 31216 33382: contig of 2167 bp in length  
\* 33383 36316: contig of 2934 bp in length  
\* 36317 40441: contig of 4125 bp in length  
\* 40442 44615: contig of 4174 bp in length  
\* 44616 49292: contig of 4677 bp in length  
\* 49293 53647: contig of 4355 bp in length  
\* 53648 59590: contig of 5943 bp in length  
\* 59591 61868: contig of 2278 bp in length  
\* 61869 67152: contig of 5284 bp in length  
\* 67153 71513: contig of 4360 bp in length  
\* 71513 76928: contig of 5416 bp in length  
\* 76929 82257: contig of 5329 bp in length  
\* 82258 89418: contig of 7161 bp in length  
\* 89419 103318: contig of 13900 bp in length  
\* 103319 113206: contig of 9888 bp in length  
\* 113207 133155: contig of 19949 bp in length.

## FEATURES

source

1. .133155  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/clone="RPC198-2017"

BASE COUNT 38122 a 28987 c 27853 g 38141 t 52 others  
ORIGIN

## Alignment Scores:

Pred. No.: 2,44e+03 Length: 133155  
Score: 52.00 Matches: 11  
Percent Similarity: 65.22% Conservative: 4  
Best Local Similarity: 47.83% Mismatches: 8  
Query Match: 40.31% Indels: 0  
DB: 2 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x AC011910 (1-133155)

QY 4 LeuGluaLeuAlaAspGlnThrAspAlaLeuGlnSerGluAlaAlaValLys 23

|||||

Db 99660 CTGGAAGCTCTGTTCAAGCAGGTGGAGTCTCTCCAGCGCGAGTTCCTCTGTGGCA 99719

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 98 (18), 10445-10450 (2001)  
 PUBMED 11517334  
 REFERENCE 2 (bases 1 to 17595)  
 AUTHORS Richards, S., Hekmat-Scafe, D.S., Hurd, D.D., Caprette, D.R.,  
 Saxton, W.M., Carlson, J.R. and Stern, M.J.  
 TITLE Direct Submission  
 JOURNAL Submitted (06-OCT-1998) Biochemistry and Cell Biology, Mail Stop  
 140, William Marsh Rice University, 6100 South Main, Houston, TX  
 77005, USA

FEATURES  
 source Location/Qualifiers  
 1..17595  
 /organism="Drosophila melanogaster"  
 /db\_xref="taxon:7227"  
 /chromosome="2"  
 /map="28E1"  
 360  
 /note="push3420 P element insertion site"  
 repeat\_region  
 join(370)..513,695..789,913..1325,1581..3747,3808..4335,  
 4404..10389,10447..10651,10701..11393,11455..13118,  
 13175..13657,13715..15650,15706..17302,17366..17584)  
 /gene="push"  
 /product="pushover"  
 370..17584  
 /gene="push"  
 /gene="push"  
 join(395)..513,695..789,913..1325,1581..3747,3808..4335,  
 4404..10389,10447..10651,10701..11393,11455..13118,  
 13175..13657,15706..15715)  
 /gene="push"  
 /codon\_start=1  
 /product="pushover"  
 /protein\_id="AAD20449.1"  
 /db\_xref="GI:4426610"  
 /translation="MSAHSQGTOWNSVVKALILNRGTALNKNVNEVLLKAITCEHDF  
 FEENSTQFTVTAALAAADKLMQIKTICOTICQLHDATAVLIRFIYLRPVSYVE  
 TKWLGLAKMLCEGREGCPASASSMDYNAVNLKCKHPSTTKIMPSSSSGSG  
 ASNDKSPKEIRSRSDLSVILQQLIAPLEPGKWTWPLSEVIDCTQILAAWVE  
 YFQNGVDLLDVCPLINRIRSKIMETTINGKRSYLPLTQVEATAYATNSHML  
 TDTILSQAALLIEMOPLTNSRISMCGIAALYNAVLTSIATSVLGMQASSSQKQ  
 TASTSQSGVSGGQSNRHDHDFDQACSIVNKALEIYNSIGHMFKTSARIRHVQK  
 HLCYGWELLISGQAGSAGSGSSDSKASKATKSGSEAGTAPTPPIARVNLF  
 KVQGFGLAAATANHSIKLSELIEDLVKAAAGQSLESTELPEPAQDPIRLYS  
 ERIVRLVNTLHQLFTFLAVAYRKACTLKASAKDRTECEPISYSDSTYRNDLS  
 CDSNEDDESILGHWFETLSPETHDDNANTSTQERAQKQSAVLPKDEPHYLDL  
 AADIFCDQFLANRHMORYKVGVSQDOOMLMANIKDFDVRNRETDOGSGNA  
 PAASAGATAGASTKQWTSIRFSGAAGYIHNLTSLSEQLQSLQHLHSISPW  
 STNTNTPVLAQYPTSLVQIILLKPTQKEACLSVWHRLINTLVEGVCSNTASD  
 SYEDLNIIEAHLMLFHLNLMQKSIILLTAGVIRCAEYVCRGISEDRPVKNSQI  
 MLISRLALEYLMKHLYNAPPELLDOVRNLEFSVSSMPDQKITDILNCRTKLNSYC  
 RHDIENFRKAGEYSSIRPTYSYLVAGDPETSYAQAEKLDGLAWNFIICLPDKLK  
 YPLIVDALDILSLTDSMSYSKEDKESMHNLCIOYCTIAWKLNLGLPPSTSHVE  
 SLKAERPNLHLSMISIRUPLASHYLVSSLIKQGNITQYAEFLTHVGDIGADIKY  
 SLKOTILGVEAFNSQNGNTPRLSDILFDLSVAHMQAVAMANKGLKWPRESIDA  
 AGEQACTSLTSSNDPELYSSNESIDQKLKDDCKLSDLKQYNOLYNELMVKLM  
 LKRVAYVMQROFFKQDQSGSLTLLSFTGSLPSVYARKMLQFVSNRLFLQSLQXK  
 QFHEDIVECFSELATVDVARLQWLGHIYIGPNVYDSTSEALDVTCTGMTLSILOP  
 SSSSSSNAOFTPNMATYAMPSTISDQDPMEIEYDCCTAAGGSEPNCTGGAANTSQ  
 ILSLWAKAOPDPSESSQACDHSERNGALLSFYKSLVKDQSKASQIAPPLFOALL  
 QLGQTLSPQEGCDFADVLQIMITLADAPARGHVALFNTLILLELAKLQLPDKHL  
 KHAENVSAQURLYSELISGIFGRSQHNPDPDELDQTDIDELYDELAEEGQDLSLD  
 DNSDDETLNKLCTFSQKQFMQHWYHCHTCNNINTVGVCSVCARCKHGHVSYAK  
 YGNPFCDCGAKEDSCALLSRIRGSNEVRDSAGISYLPSSHMLACKKRSLLPVGO  
 VLPBKDSLTERIAVLTKLEPYRETIQHODQWLLVVRCLLEVFDLILLSTINENCLY  
 SVLCCHRRATAALERHLLQSGFOVTQILMPATLGSQEGAFENVRMNSGQGTQIKH  
 LLSGVRVAFPCCLSPHGRROQLAVSHEKGVITLIQLSALLQKQASAKRKLTLTOL  
 SSAPACTVITSLAANPCNEDCLAVGCKEHLTFSSSGSTNEHVYNPQLENGNYIK  
 KAPWLSQITLAVNTSDVYLYDVAVDYSPKYLYAVAGKIRDCFTFYQDGNMNL  
 TFASSGYIYTOQLDQSLVAGHGFYVNTLELSHOKIADAGHGGGVSTIYSHTLQ  
 LLFYYSYSGKSGFPLNVSQGVGIYHLDTNSASKSKSGPLQPLQWTFVYGHPL  
 VYASMQTSNPNILMITPERIYQEIKAQSAKSRIMDVVGRHVAAGVEKTYTLLLCE

variation  
 /gene="push"  
 /replace="t"  
 3600  
 /gene="push"  
 /replace="t"  
 17558..17563  
 /gene="push"  
 17584  
 /gene="push"  
 4795 a 4143 c 4251 g 4406 t

polyA\_signal  
 /gene="push"  
 17584

polyA\_site  
 /gene="push"  
 4795 a 4143 c 4251 g 4406 t

BASE COUNT 4795 a 4143 c 4251 g 4406 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 301 Length: 17595  
 Score: 52.00 Matches: 11  
 Percent Similarity: 65.22% Conservative: 4  
 Best Local Similarity: 47.83% Mismatches: 8  
 Query Match: 40.31% Indels: 0  
 DB: 3 Gaps: 0

US-09-847-539a-6\_COPY\_59\_86 (1-28) x AF096896 (1-17595)

Qy 4 LeuGluaLaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVallys 23  
 |||||  
 Db 7027 CTGAAGCTGTCTCAAGCAGGTGGAGTCTCTCCAAGCCGCGAGTTGCTCTCTTGTGCA 6968  
 |||||  
 Qy 24 AlaAspAsn 26  
 |||||  
 Db 6967 GCCACAAT 6959  
 |||||

RESULT 22  
 AC018007  
 LOCUS AC018007  
 DEFINITION Drosophila melanogaster, \*\*\* SEQUENCING IN PROGRESS \*\*\*, in ordered  
 pieces.  
 ACCESSION AC018007  
 VERSION AC018007.1 GI:6553183  
 KEYWORDS HTG; HTGS\_PHASE2.  
 SOURCE fruit fly  
 ORGANISM Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

ASNDKESPKSEIKRSRDLSSVILQLOLIAPLEPGKMTWVPLSEEVYTDCTEQIILAAVNE  
YFOEQNGVTLDDVCVSLPILNRYSKYMETINGCKSLYLPLTOVEATVAKSMNHML  
TDLRTLSQAOLIEQOPLTPRIERLSMCGTAALYNAVITSIGTSMGSAOSSQKQ  
TASTSGGSGGGSGGQKHDDFEDQACSYNKALEIYNIGHMFKTSAKHVYQ  
HLCVGSWLLSIGIAGMAGSGSGSSDSKASKATKSGSEAGTAPPTIARYNLF  
KVQGFGLNAAIHNSIKLSLEITDLKVAACQSLSELEPEAPQDILQNTYSSL  
ERIVRLNATLHOLFTLAVYAKACTLRASAKDRTECEPSYSOSTYFNDSLS  
CDSNDESSSYLGHWFKEITLSPETHDDNANTSTQERAEOKSALVPLDRPHELDL  
AADLFCFLDFLANRHAYQRYVYKAGVSQOQMLMANILKDFDRVMNENYDQGSNA  
PAASAGATSGAGTQWTSIMRFSGAAGRYIHNLSITSESEQNLQHLISLSPW  
STDNTPWQYPSYLSVLQILKLTQKEAKACLSYWHRLINTLVGVGCSNSTASD  
SYEDINIEHAQLLFLPHSLNLMOKKISILLTAGVYIRCAEVCYKISDPRVKNQSI  
MYSLLLEFLHKLHLYNAPPELLDVWRNLFVSSMPDQKITDILNCRVTKNSYC  
RHDIEENFKSAGEVSSIRTEYSLVMDPEISYMAOEFLDGLAWNFIILCTPDKLK  
YPLVADLTILSITDSMWYSEKDKKEASMINLCAIQCYTIANKLWILPPSTSHVE  
SLKASPNLHMSIYIPLASSHYLVVSLIKOQYTOYAEATLWTHVGDIGADIKY  
SLKOTILGEAFNSQMGNGTPIRSLDILFDLSVAHQAVAWANKEGYLNQVRKRESEDA  
AQEGAGTSLTSSSDNPELYSNESIDQKQDDGKLSLDQKLYNOLVNLVYKLM  
NSYQLLSIVRGOMLKSSSTPEKALNLIPIVSDKPAIMLEHAALFKLLPNEDKO  
LIANWPKCLVMDFAFKGKHOPVPEYILNVIDAHI IELTRGSNTSLHTLKHCLKSI  
LOPELLLPHTANADEVOLKOLLSIKMLDRTDYLOGHSEHCLREILSLGTQBAOK  
LLEBYHYGYCIRMLKEFAADLRQPAAGPGSGAPLQDQDRAFMNESMLFAVLKTMFKM  
LEKPVAVAOXOFFKQDRSGSLTLLLSFTGTSPLSVYARKMLQFVNLRLTQSLQADS  
QFQHDLYVECESELATVDVARKLQWILHI IYGPVNSTDVTSEALDITTCRLMSTLQP  
SSSSSNAQPTNATVSAMPSISQLDPMETIEYDCGTAAAGGESNPGTGGAAANTSO  
ILSLWQAAQPPDSESSQADHSDSERNGALLSPVSKLVKQSKASOIAPLFOALL  
ILSLIPROEGCDFADVOIIMITLADASPARGHVALPNTLLLELAKLOPKHL  
KHAENVYSAQLYLSLQSIGFRGSRQINPFWDELQDIDELIDELAEBOQSDSLD  
DSDEDTLNNKLTFTSQKQEFNMHWCHTCNINTVGVCSVCARVCHKHGDVSYAK  
YGNFCDCAKEDSCQALSRIGNSNEVRDSAGISYLPSSHSLIAGKRSLSPVQK  
VLTRKDSITNIAVRLKLEPQYRETQLQOWLVLVRCILLEYFDLLPSINENCMILY  
STVGCKRTAALERLHLLBQSFVTDQLMFATLGSQBGFENVRMNSGDOQOTIKH  
LLSSGVVRVAFCLSPSPHRRQOLAVSHEKQVTLLOLSALLQOADSARKLTLTOL  
SSAPACTVYISLANPCNEDCLAVGLKECHILTFSSSGSTNEHIVYAPLENGNITK  
KAVPLVSGTLLVAVTSDYIKIYDIAVTYSPKYIYLVAVGKIRDTFVQDGNMIL  
TFASSGYTYTQLODQSLAVHGFYVNTLELSHOIKIDIAHIGGGVSIYYSHTLQ  
LILFYSYSGRGSFSLPTNVSEVGIYHLDNTNSASKSKAGPLOPLQVOWTEVTHPGL  
VYASMTSNPILITMTERIYLOEIKQAQSAKSRIMDVVGI RHAVAGVEKTTLLILCE  
DGLSIFSAQPEYTFMSPQVOPGNOLYSTLMKAGGSGSTSKSKSNTASGKWT  
RKASQOQKPTAGGQVPFIDFEHCLNADVEFGGNLQLIYNKQILRFLSTGMF  
DIATRNSGTLEVINNDPNVAVIGRVLIGTDQVQAPQSVITLGTPTPVRRARWF  
VUPTREMIASDKLLKVYFAPDPEHVITLDCIEVYKSKELVGMDESDVTVPG  
SSAPAVSSQSSAGNFGFCNICTQIDRMANHLEVMDCALHLLGSPVAPNRQKAVK  
TASALLLTPNPVOTOARYLATIYGTALYHNYKDGVLLOFVNRLEQSMQPKLEKL  
ETLRIDPEAYRVLVAVRGITANARPOSALKICVENNDIYPTLMGIVLELKHVPTIL  
DEPNVIRAGLCQPTIYVHCLIEIMYGFALADPGQVGRMTYFIDLLKHDAVSISHSA  
KEALTLILSPMRKRRKVAIVTPPACTPTPTSTMQALQAAASSAASDIEEAQVY  
DGSVGGGLPEPNADAEGAAYGVGGVGOQOMLNLEAFMGGFPRLGLPEADDAEAIM  
DIAIALQOHHGDANALQOGLANLOGIQTATAMAAVNAANAALSGGSDDBEIS  
NYATDGSITLRTSPAPAGSGSGSGSVESIGTTSARSSNFGDHANASPPROGSKD  
DQEQPGSVAGSGVAVLSAMSSSEDNEANEDDKLSKLHLDRITVLESITQHLGTFD  
LCNGLOAIPLOVILMLTDLNNGNERDQVLDLITLALVGLCGNKKWGAARMETKC  
PGNEVRLALLSLFVGMKTSKQGTGTTSPPHQFKNSSFVASTANYLSKSGAFVYA  
LEALNVLHMKNLVGLDGAAGLASQASGASGPGVQLLKPIKHGPKPDISIL  
IPHNLYKNYPOIDFESYDGLLIEIIVRLPYQILRLSSAHPDNYDCAMFTTICEYM  
MLNLTLLRVRKRLMYICGSEKFRMYRQGLSDAHFRVYKRVNCIVSSKTKGAPYN  
ANPNMYSDALVELTHELRTQOEISQMETGNWQKFCVVEDALAMLEIACYQLDDGV  
SPITLLOLQAVACDALIIRFVRIFFLENNITQLRWQAHSPMTGLFEAHANRQREKLLN  
CFVHQHIFRTAGTQVADLLGLYTLSTRITERLPFVSRAVDVTRQWHELICKL  
IFWMLPYLTSTPTQVAGVYLESEPCLVNCPNPEVPMANIKLPVSKSDSYVTTMYKL  
PNAPYITLSTLQVNGYVLESEPCLVNCPNPEVPMANIKLPVSKSDSYVTTMYKL  
VQETIKSLIIVRIADLKRTKMTVINVYNNRSVOAVVELKNRPAWHKARSVLSQA  
VOTELKIDPLPITACNLMEIADFETVSGSSENIQCPRCSAAVPAYPGVCGNGENV  
FOCHKRAINDKEDPFLLCHSGCKYAKDFSMYARVCCADVPIESAEEDRVKTVSIL  
HSSLERADNRTROLNKNOMEILLIKQVAHRSSDRWVEDNNASVHSTSOVNKIITLLI  
AQKYVESRTSELSKIVQKACRSELVAYDROQOQPPVPNPGSTGTAQNTNRC  
YGLKALAEQCLTLARAMAYNDCRVCLYSOGLVSELAHEHLNRRCTPLIOEVRNLLY  
VLTCDNAEACHILLQVTRVKNLWMSIPLISLEAAHQEWLLEVLIGQDDICWEY  
KLKVLFEILSNCRLPGRPVTAHLPCILRIMQNLICPVLPFGSKPQKVATVTDJCSMKM  
FEGVATVYRAMLSDNRNHEIYAAWSKRNPNQAKLKNADQNVASGSDAPKSSRRE  
VGNATVLSKYGKWRNREVDKQVRIKPLVNAKIQPLLFNANSRFGQLASLSSSL  
SRTNRRQOALNMLTSLFKHVGEAGEAAYELMLYKNMATQPMQVLYLVKGLVSOIS  
QLLAIEISKVHRMBEYLSLSDLSGLYALRQYVELLWLLLECPNIRRTYKTRMLPVLE

SYLAURLSVVORTLIDDQAEKLEMLEMDMTSGTBEETRAFMEILLIDTVEKTRMNDIK  
TPVVFPERLYSIIHPDEHDESEFYMTLEKDPQOEDFLOGRMIGNPNYPSPSEMGIGPLMR  
YMKRLCTDCEALLEEDDNGHLLVNNKIIISLDLPVKDYKKYKWLAEGGDORAMRIY  
YMRGLLDGATEEFVETLNNKSOEQVDTEOLYRMANVLADCNGLRMLERIGISQARIS  
NRRELIQTLKFLICVVRQCOELVCPGEIGAINTLIKVLOMQCLOSENDISQSAVTE  
QLLEIMETILSKAASDTLDSFLQSLTFFGPEYVALLISCTDCPNVRNPNVSLRHILIR  
VLAALVYGENVKMALLCHFKDTLNFNRDNERTPEEFKLELPCVLTNQIERNCTGG  
TLKDYIVSIVGIVEKMAIYTERHAPCVKPTLLRTDDELKEFISRPSLKYILRPLTGLS  
NHEATQVAISKDIIPIIHRLEQVSSDEHVGSLAENLLEALSTDSATARVQOVRDFT  
RAEKKRLAMATREKQDADALGMRTNEKQVYTAGSILQTEKURDETGLTCTFCREGYA  
COPDKVLGIYTTKRCNVBEFELKSKTIGYTTVTHFNVVHYDCHTSAIRLTRGRDEW  
ERASIQNANTRCNGILLPLWGPAGVGAFAFSACMTRHSSYMOESTQCDISYTSVSHDLK  
LLLVLFANERSPHDDAGGQPOSNMHFVYPLLFYSVYLLSSRSAAROSKTLLTYLQAA  
PPSEKLECGEYVDCPLFNATISLSHSRELWNKHKLAHKEMIAVAGORHVSYPVLIC  
KALLAPDRQOQKTVYKPFLLMNAVLVDLILNFKTSTVSTPKEEDWPSLFDYLRKND  
EALLKSTDSIQVTLTTEEFPLCTSFVEFCQDVGALLHLEIHPDNFIEELIILAALPSTSSN  
"  
variation /gene="push"  
/replace="t"  
2672 /gene="push"  
/replace="t"  
5411.5599 /gene="push"  
/note="putative"  
/bound\_moiety="zinc"  
11402.111542 /gene="push"  
16130 /gene="push"  
3'UTR /gene="push"  
polyA\_site /gene="push"  
BASE COUNT 4288 a 3943 c 4035 g 3869 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 275 Length: 16135  
Score: 52.00 Matches: 11  
Percent Similarity: 65.22% Conservative: 4  
Best Local Similarity: 47.83% Mismatches: 8  
Query Match: 40.31% Indels: 0  
DB: 3 Gaps: 0  
US-09-847-539A-6\_COPY\_59\_86 (1-28) x AF096897 (1-16135)  
Qy 4 LeuGluaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVallys 23  
||||| ||||| ||| : : : : : ||||| : : : : : |||||  
Db 5971 CTGGAAGCTCTGTTCAAGCAGGTGGAGTCTCTCCAAAGCGCGCAGTTGCTCTCTTTGGCA 5912  
Qy 24 AlaAspAsn 26  
||||| |||||  
Db 5911 GCCGACAAT 5903  
RESULT 21  
AF096896/7c 17595 bp DNA linear INV 15-OCT-2001  
LOCUS Drosophila melanogaster pushover (push) gene, complete cds.  
DEFINITION  
ACCESSION AF096896  
VERSION AF096896.1 GI:4246608  
KEYWORDS  
SOURCE fruit fly.  
ORGANISM Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
REFERENCE 1 (bases 1 to 17595)  
Yager, J., Richards, S., Hekmat-Scafe, D.S., Hurd, D.D., Sundaresan, V.,  
Caprette, D.R., Saxton, W.M., Carlson, J.R. and Stern, M.  
Control of Drosophila perineurial glial growth by interacting  
neurotransmitter-mediated signaling pathways

```
/gene="calo"
<2..12352
/gene="calo"
/note="calossin"
/codon_start=1
/product="CALO protein"
/db_xref="GI:3893109"
/translacion="YQLLSVIRGOMLKQSSSTPEKALMLIPIVSDKPAIMLELHA
AFLLKLLPNEQKLIANPKVLNDFNFGKQHPVEPYILNVIDAHIITLGRGSLNVS
TLHLKHLKIAQILFELLHYPHTANAETOLKOLLISMLDMRTDYLOGHSEHCLR
EILSLGTOEAQKILLLEHHVGYCYMLKFERAOLROPOSAGPSGGAPLDGHRMFNE
MLFAVLMKLEKPAVQAMQKFFKQSGSLTLLLSFTGSLPVSARMLQFV
NRFLQSLQDSQFQHEHDEVECFELATVDVARKLQWGLIYIIPNVSTDSSEAL
TTRMTLSILOPSSSSSSNAQPTNMAVSAMPSISDQLDPMIEYDCGTAAGGSEN
PGTGGAANTSOILSLWQAAPNPSESSOACDHSERNGALLSFKVSLKQDQKA
SQAPLFOALLOGLTISPPOEGGDFADVOIMITLADASPARGHVALFNTLLWL
ELAKLQPKHLKHAENVSAQRLYLSLQSIGFRGRQHPWDELQIDDELDE
LAEHGEOSSLDDSDDETLNKKICTTSQOKEFMNQHWYHCHTCNMINTVGVSCVAR
VCHKGHDVSYAKGNFCDCAGKEDSCQALSRIGSNEVRDSAGIGSYLPSMSLIA
GKRRSLPQGVPLVTRKDSLTNERIAVTLKLEPYRETLOHODQWLLVVRICILEYDLA
LLPSINENCLMYSITVCGKRRATAALERLHLEOSFOVTDOLMFATLGSORGAENVRM
NYSQDQGTIKHLLSGVVRVAFCCLSPPHGRQQLAVSHEKGVTIILQSLAKQA
DASRRKTLTLLSAPACTVLSLAAPNCNEDCLAVGLKACHILTFSSSGSNEHIV
VNPQLENGTYIKKAWLPGSQTLVAVTSDYKIVDIAVDTYSPKYYILVAVGKIRDC
TFYVQDGNWMLTFASSGYITQOLQDOSLAVHGFVYVNTLBSHQHIDKIDAGHIGG
GGVSIYSHNTLQLLYFSYSGCRSFFSPLTNVSEGVKGIYHLDNNSKASAKSGPLQPL
VQTEVTGHPGLVAVSMQTSNPIILMITPERIYLOEIKAGSASRIMDVVGIRHAVA
GVKETTLLLLCEGSLRIFSAQPEYTSFMLSPOVOPGNGQLYSTLMKAGGGSGTSK
SKSNATGTSKRSKQKQKQTAGQOPVFFIDFEHCNMLADVFEGNDLLOLQINQK
KLTRFSTGMFVATNSNGFTLEVINDPNVIVIGRVLIGTQDVOVAPQSVTILGR
TIPTVFRARMFPIITREEMQLQSKLLKYFAKADPEHEVTLDDCIEVYKSGELVG
WPDESDVTPGSSAPAVSSQASANGFGFNCITQDRMANHLLVEMDCALHLLGS
GVPSMOKAVKTAASALLPTNPVQTOARYLATLYGTRALYHNVKCVLLQVYNR
ELOSMPKLEKLETRIDPEAFYRLVWGANARPOSIAKICVENNDIYPTLMG
IVLELHKVPTLEBPVNVIRGLQPETTVHCLVEIMIGFALADPGQVGRWTFIDIL
LKDHASVISHAKALILVLRPMKRRKVAIVTPPACSTPTSTSTMAQALNAASNA
SDIEEAAGVVDGSGGGGLPEPNADAGVAGVGGVQGOQMLNEAFMGGPFRL
GLPDADDEAIDATIALSQHQGDANALQSPQOGLANLQIGTQATAMAAVNAAN
VSLGSDDDGKQVQDQPSGAVSGVAVLSAMSSSEDNEADOKLSKLDLRINAV
NASPPROGTSKQDQVQPSGAVSGVAVLSAMSSSEDNEADOKLSKLDLRINAV
LESIOHLGTFDLGNGLOAIPLOVILMLTDLNGNRRQOQVLLDLHLLTADYVEIG
KVAASAREKTCFNEVREKALLSLFVGMKTKSKGTGTSPPHQKDNSSFVASTTA
NLSKSGAFYALEANTLLVHKMNLGDPYAGAGLASQASQASGASGQVQVLLK
IKHGKPFDSILIPHYKLYNPDIFESYDGLLEIIVLRYQILRLSSAHPDNYDSGF
CEAMTFTLCBYMLMLTLNRROVKLLMYICGSKFRMYRQDHSYAFVYKRVK
NIVSSKGTGAPYANPMLSYDALVELTHERLTCQETISOMRTGNWQKCVYHEDALML
METACYOLDQGVSPILLOLQAVCNLPSPGSKQAOPOSTSSASOKLHTRDEKSD
TDAYSFDPACQCTFVHQIFRACDALIIRFVIFULENNITQLRMAHSMFTGLFE
HANRQREKLLNIPWNLPLVPTVYGTQTAQFVLLGYTLSTIRLERLPEFVSAYD
VLRQNELLCNHPNAPITYTIESILOVNGYILESEPCVGNPNPEVPMANIKLP-SVKSD
SKYTTTMMIYKLVQCHTISKILVIRADLKRTKMVRTINVYNNRSVOAVKLNKRAL
WHKARSYLSQAOELKIDPLPITACNLMEIADPFETVSGSENLOQPCRSAAVPA
YXGVCNGCENFQCHKRAINYDEKDPFLCHSGCKYAKDFESMYARVCCAVDPIE
SAEDRVTVSLIHSLSERADNRYNQLTNKQMLELLQKVAEHRSDRWEDNNNAVH
STQVNNKIIOLLAQCYVESRTSPELSKIVQKACRSELVAYDROQDOPPNVPS
TGAENPTNRNVTGTCALASTEQCLTLRAMAYNDCRCLYSQGLSELAEHLNRRGT
PLIQEENRNLIVLTNDKNAACMHLLOLVTRVKNALMGSTPISLSAAVQHEMTLLE
VLLGDDICWKEKLVKIVFELFISNCRLPGRVAVLHPCLRIMONLICPVLPQSKPQ
KVATDTCMKMPEKGVYRAWLSDNRNHEYAKSMPSNNQAKLNKAKDONVARS
GGSDAPKRSREVRVAFSEKYGKRWREVRDQKRVIKPLVFNKATQPLFLNANSRP
GRQACLSLSLSTRERRQOALNMLTSFLKHGEAGEASAEYLMLYKNMATQEPNLO
YVLKGLVSQISQALATISKVRHMEYSLSLGLYALROYVELLWLLLEPNIR
TYKTRMLGPLVLSLRSVQTRILDOAEKLEMLEDMTSGTEETRAPMEILI
DTVEKTRANDIKTPVFVERLYSIIHPEHDESEFYMTLEKDPQOEDFLOAGMLNPY
PSSENGGLPBRQVKNKICDCELLALLEDNGNELLYNKKIISLDLPKDYKVKWL
AEGDRAHRIYRMRGLGDATFEFVETLNKKSQEQVDTEQLRMANVLADCNGLRV
MLERIGLSRISRRRELIQVLLKFLVCVRQCEVLOCPGEIGAINTLKLVQMCLQ
SENDSISQATEQVLELMEETLSKAASDITLDSQFSLTFEGFVYSALISCTDQPNV
RNNP-SVLRHLIRVLEALVYGENVKMALICSHFKTLNFRNFDERTPEEFKLELFCV
LTNOIEHCIGGTLKDYIVSGLIGERSLAYITTEHAPCVKPTILRTDSDDELKEFTRPS
LKYTLRFLTGLSNHHEATOVAKSDIPIIHRLEQVSSDDEHVSLSAENLEALSTDSA
TAARVQVQVDFTRAEKRLAMATREKOLDJALGMKRTNEKQVTAQKSTLQIEKURDET
```

CDS

```
GLTCTICRGYACQDPKVLGIYFTTFRKNVVEEPELKSRTIGTGTVTTHPNVHVDCHT
SAIRLTRGDEWERASLQANTRCNGLLPLWGPVAGEAAPSACMTRHSSYMOESTQRC
DISYTSVSHDLKLLVRFARFSDHDDAGGPGQSNHFPVILLFYTVYLLSSRGA
RDSKTLTYLOAPPSEKWLCEGYEDGPIFMATISLSHSLRNLNKKHLAKHLMIAV
AQGRHYSPAVICKALLAPADROVKDYTVYKPLMLMALVDLIYDNLKFTVSPKPEDM
PISLFDILKNDKALLKSDSILQITLITEELPCTSFVEFCDVAGLLHLIEHPDNFIEE
ILAAALPSTSSN"
BASE COUNT 3257 a 3028 c 3148 g 2940 t
ORIGIN
Alignment Scores: 209 Length: 12373
Pred. No.: 52.00 Matches: 11
Percent Similarity: 65.22% Conservative: 4
Best Local Similarity: 47.83% Mismatches: 8
Query Match: 40.31% Indels: 0
DB: 3 Gaps: 0
US-09-847-539a-6_COPY_59_86 (1-28) x DMEI7920 (1-12373)
Qy 4 LeuGluAlaLeuAlaAaSpGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVallys 23
|||||
Db 2329 CTGGAAGCTGCTGTTCAACGAGGTGAGTCTCTCCAAAGCCGACGTTGCTCTCTTGTGGCA 2270
|||||
Qy 24 AlaAspAsn 26
Db 2269 GCCGACAAT 2261
|||||
RESULT 20
AF096897/c
LOCUS AF096897 16135 bp mRNA linear INV 15-OCT-2001
DEFINITION Drosophila melanogaster pushover (push) mRNA, complete cds.
ACCESSION AF096897
VERSION AF096897.1 GI:4426609
KEYWORDS fruit fly.
SOURCE Drosophila melanogaster
ORGANISM Drosophila melanogaster
REFERENCE 1 (bases 1 to 16135)
AUTHORS Yager,J., Richards,S., Hekmat-Scafe,D.S., Hurd,D.D., Sundaresan,V.,
Caprette,D.R., Saxton,W.M., Carlson,W.J.R. and Stern,M.
TITLE Control of Drosophila perineurial glial growth by interacting
neurotransmitter-mediated signaling pathways
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 98 (18), 10445-10450 (2001)
PUBMED 11517334
REFERENCE 2 (bases 1 to 16135)
AUTHORS Saxton,W.M., Carlson,J.R. and Stern,M.J.
TITLE Direct Submision
JOURNAL Submitted (06-OCT-1998) Biochemistry and Cell Biology, Mail Stop
140, William Marsh Rice University, 6100 South Main, Houston, TX
77005, USA
FEATURES
source
1..16135
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/chromosome="2"
/map="28E1"
<1..16135
/gene="push"
<1..25
/gene="push"
26..15994
/gene="push"
/codon_start=1
/product="pushover"
/protein_id="AAD20450.1"
/db_xref="GI:4426611"
/translacion="MSAHSGGTWNNSVVKAILNRTGALNKNEVNLKAITRCEHDF
FEESNPTQYTAFAALAAADKLIKQIKOTICQQLDHFATVILRIFLIRPVSYVE
TKWLLGALKMLCEGPCPASSSMFDYNANVNLKSCRHPETTSIIIMPSSSSSGG
```

\* 89805 89904: gap of unknown length  
 \* 89905 95932: contig of 6028 bp in length  
 \* 95933 96032: gap of unknown length  
 \* 96033 102331: contig of 6499 bp in length  
 \* 102332 102631: gap of unknown length  
 \* 102632 111305: contig of 8674 bp in length  
 \* 111306 111405: gap of unknown length  
 \* 111406 120462: contig of 9057 bp in length  
 \* 120463 120562: gap of unknown length  
 \* 120563 130442: contig of 9680 bp in length  
 \* 130443 130342: gap of unknown length  
 \* 130343 140588: contig of 10246 bp in length  
 \* 140589 140689: gap of unknown length  
 \* 140690 149261: contig of 8573 bp in length  
 \* 149262 149361: gap of unknown length  
 \* 149362 162349: contig of 12988 bp in length  
 \* 162350 162449: gap of unknown length  
 \* 162450 173772: contig of 11323 bp in length.

## FEATURES

Location/Qualifiers  
 1..173772

/organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="RPC1-11"  
 /clone="RP11-724H13"  
 1..1282

misc\_feature /note="assembly\_name:Contig17"  
 misc\_feature /note="assembly\_name:Contig20"  
 misc\_feature /note="assembly\_name:Contig22"  
 misc\_feature /note="assembly\_name:Contig23"  
 misc\_feature /note="assembly\_name:Contig24"  
 misc\_feature /note="assembly\_name:Contig25"  
 misc\_feature /note="assembly\_name:Contig26"  
 misc\_feature /note="assembly\_name:Contig27"  
 misc\_feature /note="assembly\_name:Contig28"  
 misc\_feature /note="assembly\_name:Contig29"  
 misc\_feature /note="assembly\_name:Contig30"  
 misc\_feature /note="assembly\_name:Contig31"  
 misc\_feature /note="assembly\_name:Contig32"  
 misc\_feature /note="assembly\_name:Contig33"  
 misc\_feature /note="assembly\_name:Contig34"  
 misc\_feature /note="assembly\_name:Contig35"  
 misc\_feature /note="assembly\_name:Contig36"  
 misc\_feature /note="assembly\_name:Contig37"  
 misc\_feature /note="assembly\_name:Contig38"  
 misc\_feature /note="assembly\_name:Contig39"  
 misc\_feature /note="assembly\_name:Contig40"  
 misc\_feature /note="assembly\_name:Contig41"  
 misc\_feature /note="assembly\_name:Contig42"  
 misc\_feature /note="assembly\_name:Contig43"  
 17148..74947

misc\_feature /note="assembly\_name:Contig44"  
 75048..79238  
 misc\_feature /note="assembly\_name:Contig45"  
 79339..84833  
 misc\_feature /note="assembly\_name:Contig46"  
 84934..89804  
 misc\_feature /note="assembly\_name:Contig47"  
 89905..95932  
 misc\_feature /note="assembly\_name:Contig48"  
 96033..102531  
 misc\_feature /note="assembly\_name:Contig49"  
 102632..111305  
 misc\_feature /note="assembly\_name:Contig50"  
 111406..120462  
 misc\_feature /note="assembly\_name:Contig51"  
 120563..130242  
 misc\_feature /note="assembly\_name:Contig52"  
 130343..140588  
 misc\_feature /note="assembly\_name:Contig53"  
 140689..149261  
 misc\_feature /note="assembly\_name:Contig54"  
 149362..162349

## Alignment Scores:

Pred. No.: 2.18e+03 Length: 173772  
 Score: 53.00 Matches: 11  
 Percent Similarity: 77.78% Conservative: 3  
 Best Local Similarity: 61.11% Mismatches: 4  
 Query Match: 41.09% Indels: 0  
 DB: 2 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x AC087862 (1-173772)

QY 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaLaval 21

Db 173045 CTGGAGGCTCTGCTGCACAGACAGCGGCGACAGCAAGGAGTAGCCTTG 172992

## RESULT 19

DME17920/c

LOCUS Drosophila melanogaster mRNA for callosin protein.  
 DEFINITION  
 ACCESSION Y17920  
 VERSION Y17920.1 GI:3893108

KEYWORDS calo gene.

SOURCE fruit fly.

## ORGANISM

Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

## REFERENCE

1 (bases 1 to 12373)  
 Xu,X.Z., Wes,P.D., Chen,H., Li,H.S., Yu,M., Morgan,S., Liu,Y. and

## AUTHORS

Montell,C.

Retinal targets for calmodulin include proteins implicated in

synaptic transmission

J. Biol. Chem. 273 (47), 31297-31307 (1998)

## JOURNAL

MEDLINE

99030403

## REFERENCE

2 (bases 1 to 12373)

## AUTHORS

Montell,C.

## TITLE

Direct Submission

## JOURNAL

Submitted (11-AUG-1998) C. Montell, The Johns Hopkins University

School of Medicine, Department of Biological Chemistry, 725 N.

Wolfe Street, Baltimore, MD 21205-2185, USA

## FEATURES

Location/Qualifiers

1..12373

/organism="Drosophila melanogaster"

/db\_xref="taxon:7227"

1..3316

/organism="Drosophila melanogaster"

/db\_xref="taxon:7227"

/map="28E1-2"

/tissue\_type="retinal"

/dev\_stage="adult"

2..12352

gene

```
repeat_region 31133..31165
/rpt_family="ERVK"
repeat_region 31172..31597
/rpt_family="ERVK"
repeat_region 31598..32237
/rpt_family="ERV1"
repeat_region 34473..35285
/rpt_family="L1"
repeat_region 35883..36498
/rpt_family="ERV1"
repeat_region 38463..38929
/rpt_family="MERL_type"
repeat_region 39354..39609
/rpt_family="L1"
repeat_region 42594..42877
/rpt_family="Alu"
repeat_region 43519..43817
/rpt_family="Alu"
misc_feature 45998..46226
/note="similar to EST AA909108 (NID:g3048513) om32d11.s1"
repeat_region 47950..48034
/rpt_family="L1"
repeat_region 48869..48985
/rpt_family="Alu"

Alignment Scores:
Pred. No.: 1.45e+03 Length: 117309
Score: 53.00 Matches: 11
Percent Similarity: 77.78% Conservative: 3
Best Local Similarity: 61.11% Mismatches: 4
Query Match: 41.09% Indels: 0
DB: 9 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x AC007274 (1-117309)

Qy 4 LeuGluaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaVal 21
|||||
Db 96844 CTGGAGGCTCTGCTGGACAGGACAGACGGCGGACAGAGCAAGGAAGTAGCCTTG 96897

RESULT 18
AC087862/c 173772 bp DNA linear HTG 19-FEB-2001
LOCUS AC087862
DEFINITION Homo sapiens chromosome RPC1-11 clone RP11-724H13, WORKING DRAFT
ACCESSION AC087862
VERSION AC087862.3 GI:12963066
KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
WATERSTON, R.H.
The sequence of Homo sapiens clone
Unpublished
2 (bases 1 to 173772)
WATERSTON, R.H.
Direct Submission
Submitted (02-FEB-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Feb 18, 2001 this sequence version replaced gi:12745114.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0724H13
----- Summary Statistics -----
Sequencing vector: M13, 100%
Sequencing vector: plasmid; 0%
Chemistry: Dye-primer ET; 100% of reads
```

```
Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 150537 bases at least Q40
Consensus quality: 158418 bases at least Q30
Consensus quality: 162327 bases at least Q20
Insert size: 191000; agarose-fp
Quality coverage: 2.88 in Q20 bases; agarose-fp
Quality coverage: 3.33 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 37 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1282: contig of 1282 bp in length
* 1283 1382: gap of unknown length
* 1383 2426: contig of 1044 bp in length
* 2427 2526: gap of unknown length
* 2527 3572: contig of 1046 bp in length
* 3573 3672: gap of unknown length
* 3673 6159: contig of 2487 bp in length
* 6160 6259: gap of unknown length
* 6260 8183: contig of 1924 bp in length
* 8184 8284: gap of unknown length
* 8284 11014: contig of 2731 bp in length
* 11015 11114: gap of unknown length
* 11115 12517: contig of 1403 bp in length
* 12518 12617: gap of unknown length
* 12618 14315: contig of 1698 bp in length
* 14316 14415: gap of unknown length
* 14416 17095: contig of 2680 bp in length
* 17096 17195: gap of unknown length
* 17196 19263: contig of 2068 bp in length
* 19264 19363: gap of unknown length
* 19364 22747: contig of 3383 bp in length
* 22747 22846: gap of unknown length
* 22847 24581: contig of 1734 bp in length
* 24581 24680: gap of unknown length
* 24681 27136: contig of 2456 bp in length
* 27137 27236: gap of unknown length
* 27237 30151: contig of 2915 bp in length
* 30152 30252: gap of unknown length
* 30252 33090: contig of 2839 bp in length
* 33091 33190: gap of unknown length
* 33191 36912: contig of 3722 bp in length
* 36913 37012: gap of unknown length
* 37013 40616: contig of 3604 bp in length
* 40617 40716: gap of unknown length
* 40717 44847: contig of 4131 bp in length
* 44848 44947: gap of unknown length
* 44948 50238: contig of 5291 bp in length
* 50239 50338: gap of unknown length
* 50339 53196: contig of 2858 bp in length
* 53197 53296: gap of unknown length
* 53297 57538: contig of 4242 bp in length
* 57539 62554: gap of unknown length
* 62555 62654: gap of unknown length
* 62655 66368: contig of 3714 bp in length
* 66369 66469: gap of unknown length
* 66469 71647: contig of 5179 bp in length
* 71648 71747: gap of unknown length
* 71748 74947: contig of 3200 bp in length
* 74948 75048: gap of unknown length
* 75048 79238: contig of 4191 bp in length
* 79239 79339: gap of unknown length
* 79339 84833: contig of 5495 bp in length
* 84833 84933: gap of unknown length
* 84934 89804: contig of 4871 bp in length
```







```
CDS
5899..7500
/function="putative membrane protein"
/notes="ORF6"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAC93954.1"
/db_xref="GI:17974219"
/translation="MTLTPLSKORPGTPSPRASDAPPTASATTAASPAAPA
PETATAGTGRWSLPGGLTQAQATAGAIUALALAGIGLYLSFEHVATFAHOKL
GFATLGRKLTGVGVGILVLIADLLMAWRPIGWVRFVWLLTGATIVILNAASA
APGAWTALVYATHAIPVLFIAIVEVGKTAYDVRVPDAHNGPGVRLRWFLA
PAPTFALRMKLMWATVAEAVELDKELRVYRMLEKREYGSVKAPSDVRLPLTWER
YGLTIAELALPQAEQAOLREAEEDARLAEARKOERAAAAETARLKAACAGVAA
RTEVNSTSPAEVQAAAEVAAERAAETAEVQSEAAAAEAEDRAAEHRAEARRR
AACADAEEAREAREAAALTAEDARTAAEAESTEERARASQAORDAAEAQAAA
EAHRAAEAREAREAEDELTLSPRKVKRVKVMILAEAGHPENLPLESVMTACS
SQTASEYRSAAQLLAEGYTPAR"
BASE COUNT 1125 a 2999 c 2583 g 979 t
ORIGIN

Alignment Scores:
Pred. No.: 86.7 Length: 7686
Score: 53.00 Matches: 13
Percent Similarity: 65.38% Conservative: 4
Best Local Similarity: 50.00% Mismatches: 9
Query Match: 41.09% Indels: 0
DB: 1 Gaps: 0

US-09-847-539a-6_COPY_59_86 (1-28) x SC0414671 (1-7686)
Qy 3 AlaleuGlualaleuAlaAspGlnThrAspAlaleuGlnSerGluAlaAlaAlaVal 22
Db 6964 GCCAGCGCGCGGACCGAGCGAGCGCGTCCAGTCGGAGCGCGCGCGGCC 7023
Qy 23 LysAlaAspAsnAlaAla 28
Db 7024 GAACCGCGCGCGCGGCC 7041

RESULT 16
SCSCP2/c
LOCUS
DEFINITION Streptomyces coelicolor plasmid SCP2.
ACCESSION AL645771
VERSION AL645771.1 GI:17026158
KEYWORDS bldA regulation; GntR-family regulator; IS1648; partitioning protein ParA; plasmid transfer; Tn5714; TraA; TraB; transposase; TraX.

SOURCE
ORGANISM Streptomyces coelicolor.
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
1 (bases 1 to 31317)
Kieser, T., Bibb, M.J., Buttner, M.J., Chater, K.F. and Hopwood, D.A.
Hopwood, D.A. (Eds.);
PRACTICAL STREPTOMYCETES GENETICS: 1-1;
John Innes Foundation (2001)
ISBN 0-7084-0623-8.
2 (bases 1 to 31317)
Haug, I., Brolle, D., Weissenborn, A., Kieser, T., Wohlleben, W. and Altenbuchner, J.
Unpublished
3 (bases 1 to 31317)
Bentley, S.D., Parkhill, J., Barrell, B.G. and Rajandream, M.A.
Direct Submission
Submitted (14-NOV-2001) Streptomyces coelicolor sequencing project, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
CB10 ISA E-mail: barrell@sanger.ac.uk Plasmid supplied by Prof. David A. Hopwood, [3] John Innes Centre, Norwich Research Park, Colney, Norwich, Norfolk NR4 7UH, UK
FEATURES
source 1..31317
/organism="Streptomyces coelicolor"
/db_xref="taxon:1902"
1..31317
/organism="Streptomyces coelicolor A3(2)"
/db_xref="A3(2)"
/db_xref="taxon:100226"
/clone="plasmid SCP2"
425..1402
/notes="Insertion element IS1648 bounded by 32bp imperfect inverted repeat. Copies of this IS element are also present in the Streptomyces coelicolor chromosome (EMBL:AL023861 and AL031035)."
492..917
/gene="SCP2.01"
492..917
/gene="SCP2.01"
/notes="Identical to Streptomyces coelicolor putative IS1648 transposase SC3C8.15 or SC6A9.36 TR:069927 (EMBL:AL023861) (141 aa) fasta scores: E(): 0, 100.0% id in 141 aa. Contains Pfam match to entry PF01511 Transposase_6, Transposase. Complete protein may be translated via a frameshift into the downstream CDS."
/codon_start=1
/transl_table=11
/product="putative IS1648 transposase"
/protein_id="CAD11995.1"
/db_xref="GI:17026159"
/translation="MGRGLTNRWSLLEPHLPGLGGRGRRNDHRTVNGILFRVRT
GVPRDLPERYGSWKTYERHRRMSADGTWDRILQSVQADADLAGRIDWSMVGVDSTS
CRAHOHAAGARKTRPVKPKQRPQTAPTTRDSGPGCAA"
594..839
/gene="SCP2.01"
/notes="Pfam match to entry PF01511 Transposase_6, Transposase, score 12.30, E-value 3.6e-05"
914..1372
/gene="SCP2.02"
914..1372
/gene="SCP2.02"
/notes="Identical to Streptomyces coelicolor putative IS1648 transposase SC3C8.16 or SC6A9.37 TR:069928 (EMBL:AL023861) (152 aa) fasta scores: E(): 0, 99.3% id in 152 aa. Complete protein may be translated via a frameshift from the upstream CDS."
/codon_start=1
/transl_table=11
/product="putative IS1648 transposase"
/protein_id="CAD11996.1"
/db_xref="GI:17026160"
/translation="MTCKIHLAGEGGCRPLALLTPGOWGDAPOLVFLDRIRVPRPL
GGPRTRPDHVGDKAYSSRRNRYLRRIIRITIPKQQRANRRRGEGGCPAGF
DRDHYRRNVEVERTINRLKNPRAVATRYDKRAYVFGTGTAAALRLWLQ"
complement(3033..4010)
/gene="SCP2.04c"
complement(3033..4010)
/gene="SCP2.04c"
/notes="Unknown"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAD11997.1"
/db_xref="GI:17026161"
/translation="MSRRSLALPSTREFDDADELEAAPEEKLAARSAGVVASLTG
ADLSPLTVAQLTPPYDAETVAPLNDQERYLDVCEQALHGFRKSVVAGKALEVI
NRGLRYRTHETTFADYVTEYMDKRAHAYRMIEGRPADLYSPGIDNEQAGALEPV
LKEYGPEVTYLRGVKELRGDRRVTAADLSKRAALPPPKHRLQAPDRDLTVAAA
EGRAPLAPAEKPVPAQAADEHOAEQVDEGGVSDQVDEGAETATLECAVAQORQIV
DRVGSGTTLAALLYDPCRGDHLRELQYQAQYRAYRARDTSGEQVADDA"
complement(4007..4939)
/gene="para"
/notes="SCP2.05c"
complement(4007..4939)
/gene="para"
/notes="Similar to Caulobacter crescentus chromosome
```

mrna	join(7165..7303,8920..9534,10965..>11147) /gene="CG7901" /product="CT42545" /db_xref="FLYBASE:FBan0007901" /db_xref="FLYBASE:FBgn0038572" 7165..>11147 /gene="CG7901" /map="90E3-90E5" /db_xref="FLYBASE:FBan0007901" /db_xref="FLYBASE:FBgn0038572" join(9142..9534,10965..11147) /gene="CG7901" /note="CG7901 gene product [alt 1]" /codon_start=1 /db_xref="FLYBASE:FBan0007901" /db_xref="FLYBASE:FBgn0038572" /protein_id="AAF55499.1" /db_xref="GI:7300340" /translation="MDNEALDPTIKSTSAATPTAAASETTTAASSVWETTTIAAA TASAAESKNETTANNNSNTSGSISSNNNIVIPASATNGIKESNLSNTTTTAAAV AAATVEGVAITSTIVTGGTPTLSSLANKLKDNTPPYDAPPTPISKVLNITGTP IVRKEKQTSARYNASKNCELTALIPLESE" join(9142..9534,10965..11147) /gene="CG7901" /note="CG7901 gene product [alt 2]" /codon_start=1 /db_xref="FLYBASE:FBan0007901" /db_xref="FLYBASE:FBgn0038572" /protein_id="AAF55500.1" /db_xref="GI:7300341" /translation="MDNEALDPTIKSTSAATPTAAASETTTAASSVWETTTIAAA TASAAESKNETTANNNSNTSGSISSNNNIVIPASATNGIKESNLSNTTTTAAAV AAATVEGVAITSTIVTGGTPTLSSLANKLKDNTPPYDAPPTPISKVLNITGTP IVRKEKQTSARYNASKNCELTALIPLESE" join(12280..12385,12562..13626,13976..14480,14769..14965, 15039..15373,15432..16135,16241..16677) /gene="CG7913" /product="CT23900"	JOURNAL REFERENCE AUTHORS TITLE JOURNAL UNITED KINGDOM FEATURES source 1..7686 /organism="Streptomyces coelicolor" /plasmid="2 SCP2" /db_xref="taxon:1902" 1261..1713 /function="unknown" /note="ORF1" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="CAC93949.1" /db_xref="GI:17974214" /translation="MSARPRTPHOFDGGYTLQAVGSDVPLYSLGAAAEWATLEWLR EHGGAAGYVPTPEELGEDYGCASKDTCRKALNRLVKGVLVVKPGPRSGSYQLNPLRYW EGAGSTOVNACRRMAPPRVAPDDRAMTRSAKPKTIPATRRRAAGETR" 1710..2129 /function="unknown" /note="ORF2" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="CAC93950.1" /db_xref="GI:17974215" /translation="MTTVPVGFNPEDLTAPSLYSLNLSAAQHCTLAWEDHGGLFD VIVPVLTVAEDCGNSVSTVHEALARLEALNLVLTSGALYRINARVYFTLHPLELM ITAAITDPPVTPDORARAPRKVSNTDARRRRTIRPV" 2140..2179 complement(2884..3219) /function="unknown" /note="ORF3" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="CAC93951.1" /db_xref="GI:17974216" /translation="MTNELDTASAAQETLERSAAELVRLQVVRELRAHGLVYRALD GMEDPYLVAEGREFVHRHGPAYLHSWVLYPYGDODEPTAEAPDFNGPFGTARAA AASLKPGQA" complement(2884..3219) complement(3216..4013) /function="putative transcriptional regulator" /note="ORF4" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="CAC93952.1" /db_xref="GI:17974217" /translation="MATRRKGRPGGYEIAAHFRRLMDSGELSPGDPLPSMRDYCDQF GSAITTVNRAFLQEEERTVSKPGVTIVRDMSKRVVPFSTYGDVLAPGDKGPWER ATAAGLDGRMLVEAPEVGAPADVAARLGEALGVVHRRRATIGEDVVLODQAWY PLETARAAGLDPRKGVGVLGAMTGAGLSPTSDHDEVEVWVPSAQOAELSLGRVS VLVVERVYDTRVRLTRITGAADRLITLYKGLPLKATGAGSTS" 4425..4904 /function="unknown" /note="ORF5" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="CAC93953.1" /db_xref="GI:17974218" /translation="MIRSVTTRARAFRTVADVRSTVAPASLVRSRADETAPDPTGL LAPEIEARAEIEADAADYAAATDRASGRGKRKARKLLDRLPACRGYGVMTVERVP SSRQTPDLEAIRATYKRLGLGDVPMRTCAPSLRVTRAADVVPFADEATAQLAA" 4924..4947	actinomycetes J. Mol. Microbiol. Biotechnol. In press 2 (bases 1 to 7686) Wilkinson,C.J. Direct Submission Submitted (28-SEP-2001) Wilkinson C.J., Department of Biochemistry, University of Cambridge, 80 Tennis Court Rd., Cambridge, CB2 1GA, UNITED KINGDOM
gene			
CDS			
terminator			

Brandon, R.C., Rogers, Y.H., Blazej, R.G., Champe, M., Pfeiffer, B.D., Wan, K.H., Doyle, C., Baxter, E.G., Helt, G., Nelson, C.R., Gabor Miklos-G.L., Abrell, J.P., Agbayani, A., An, H.J., Basu, A., Andrews-Pfannkuch, C., Baldwin, D., Ballew, R.M., Beeson, K.Y., Baxendale, J., Bayraktaroglu, L., Beasley, E.M., Beeson, K.Y., Benos, P.V., Berman, B.P., Bhandari, D., Bolshakov, S., Borikova, D., Botchan, M.R., Bouck, J., Brokstein, P., Brottier, P., Burtis, K.C., Busam, D.A., Butler, H., Cadieu, E., Center, A., Chandra, I., Cherry, J.M., Cawley, S., Dahlke, C., Davenport, L.B., Davies, P., de Pablos, B., Delcher, A., Deng, Z., Mays, A.D., Dew, I., Dietz, S.M., Dodson, K., Doup, L.E., Downes, M., Dugan-Rocha, S., Dunkov, B.C., Dunn, P., Durbin, K.J., Evangelista, C.C., Ferraz, C., Ferriera, S., Fleischmann, W., Foster, C., Gabriellian, A.E., Garq, N.S., Gelbart, W.M., Glasser, K., Glodek, A., Gong, F., Gorrell, J.H., Gu, Z., Guan, P., Harris, M., Harris, N.L., Harvey, D., Helman, T.J., Hernandez, J.R., Houck, J., Hostin, D., Houston, K.A., Howland, T.J., Wei, M.H., Ibegwam, C., Jalali, M., Kalush, F., Karpen, G.H., Ke, Z., Kennison, J.A., Ketchum, K.A., Kimmel, B.E., Kodira, C.D., Kraft, C., Kravitz, S., Kulp, D., Lai, Z., Lasko, P., Lei, Y., Levitsky, A.A., Li, J., Li, Z., Liang, Y., Lin, X., Liu, X., Mattel, B., McIntosh, T.C., McLeod, M.P., McPherson, D., Merkulov, G., Milshina, N.V., Mobarry, C., Morris, J., Moshrefi, A., Mount, S.M., Moy, M., Murphy, B., Murphy, L., Muzny, D.M., Nelson, D.L., Nelson, D.R., Nelson, K.A., Nixon, K., Nusskern, D.R., Pacle, J.M., Palazzolo, M., Pittman, G.S., Pan, S., Pollard, J., Puri, V., Reese, M.G., Reinert, K., Remington, K., Saunders, R.D., Scheeler, F., Shen, H., Shue, B.C., Siden-Kianos, I., Simpson, M., Skupski, M.P., Smith, T., Spier, E., Spradling, A.C., Stapleton, M., Strong, R., Sun, E., Svirskas, R., Tector, C., Turner, R., Venter, E., Wang, A.H., Wang, X., Wang, Z.Y., Wassarman, D.A., Weinstein, G.M., Weissbach, J., Williams, S.M., Woodage, T., Worley, K.C., Wu, D., Yang, S., Yao, Q.A., Ye, J., Yeh, R.F., Zaveri, J.S., Zhan, M., Zhang, G., Zhao, O., Zheng, L., Zheng, X.H., Zhong, F.N., Zhong, W., Zhou, X., Zhu, S., Zhu, X., Smith, H.O., Gibbs, R.A., Myers, E.W., Rubin, G.M. and Venter, J.C.  
The genome sequence of *Drosophila melanogaster*  
Science 287 (5461), 2185-2195 (2000)

2 (bases 1 to 224896)  
Adams, M.D., Celniker, S.E., Gibbs, R.A., Rubln, G.M. and Venter, C.J.  
Direct Submission  
Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive, Rockville, MD, USA

FEATURES  
source  
1..224896  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/chromosome="3R"  
complement(join(1202..2465,2540..3150,3540..3806))  
/gene="CG7218"  
/product="Ct2261"  
/db\_xref="FLYBASE:FBan0007218"  
/db\_xref="FLYBASE:FBgn0038569"  
complement(<1202..>3806)  
/gene="CG7218"  
/map="90E2-90E4"  
/db\_xref="FLYBASE:FBan0007218"  
/db\_xref="FLYBASE:FBgn0038569"  
complement(join(1245..2465,2540..3150,3540..3738))  
/gene="CG7218"  
/note="CG7218 gene product"  
/codon\_start=1  
/db\_xref="FLYBASE:FBan0007218"  
/db\_xref="FLYBASE:FBgn0038569"  
/protein\_id="AAF55495.2"  
/db\_xref="GI:10726596"  
/translation="MNATLNSAGKKRQLRFRGDVTSRVEELHQQOQBOKAPLAQ  
DIVAATPATPAQAQROLOSCTACTNTYDFKQVEMTGYMLEHDEERYSARQ  
KIYSPMRPRDLERFWYVIMQCADSEFLYIHFTLPVRFVMAVYLSRTVARIPLRS  
SGRLLSFAEICDLKGVIMVTITMLLDVTRNHYIIKSSQSIKLYIFYINWLEVD  
RLLSAFQGDITDALFTATPEKNSRHEFGVLTVLTIYFLHSLGIMFQATCLNY  
AVNSNKGILLTITNNFVELGSKVKKFKNLQFQTCSDRFRHLSLVLFVIVIO  
TKKEFDWISITQCVMLPCFVAVLEITLIDWKHAFITRFNLPESIVREYTTSLAYD  
MTQTRQKHAFSDHSLVARRMGFIFFPLAVVLIKALITYAVSFENLAANLLFLAYLFA

mRNA  
MGLRICLTICALGCKACKLMKEHOTERNSTPSSMTNPVIGAAAPYSAATGQGNHNN  
NNNNNSISGSKPAQVTTLLTPPSAGHLDVSKNFSRTSIASSTSPKKAVSQEILDV  
TNSLEUGATATLSNSDLDVCLNEQVNTNSAVQYIQODLVQSRQPDMLLNN  
SGDVTSAKAKKATORLPKTRKSEFSPMSVKEGGAAGIAGGNQTTQL  
join(4128..4212,4268..4485,4610..4873)  
/gene="Cbp20"  
/product="Ct23846"  
/db\_xref="FLYBASE:FBan0012357"  
/db\_xref="FLYBASE:FBgn0022943"  
complement(<4128..>4873)  
/gene="Cbp20"  
/note="CG12357"  
/map="90E3-90E4"  
/db\_xref="FLYBASE:FBan0012357"  
/db\_xref="FLYBASE:FBgn0022943"  
join(4165..4212,4268..4485,4610..4808)  
/gene="Cbp20"  
/note="Cbp20 gene product"  
/codon\_start=1  
/db\_xref="FLYBASE:FBan0012357"  
/db\_xref="FLYBASE:FBgn0022943"  
/protein\_id="AAF55496.1"  
/db\_xref="GI:7300337"  
/translation="MPASVELSSYRDQHFKEGSRSEQSLRDSCTLYVGNLSFYTTTEE  
OHELFSRCGDVRIYVMGLDKYKTKPCGCFEYVYRSEAAAMRVNCTRLDRLIR  
VMDAGFVSGRQYGRCKTGQYRDEYRTDYDAGRGYKLLSOKIAPNTDR"  
complement(join(4878..5331,5462..5596,5791..5935))  
/gene="CG7217"  
/product="Ct22257"  
/db\_xref="FLYBASE:FBan0007217"  
/db\_xref="FLYBASE:FBgn0038570"  
complement(<4878..>5935)  
/gene="CG7217"  
/map="90E3-90E4"  
/db\_xref="FLYBASE:FBan0007217"  
/db\_xref="FLYBASE:FBgn0038570"  
complement(join(5002..5331,5462..5596,5791..5799))  
/gene="CG7217"  
/note="CG7217 gene product"  
/codon\_start=1  
/db\_xref="FLYBASE:FBan0007217"  
/db\_xref="FLYBASE:FBgn0038570"  
/protein\_id="AAF55497.1"  
/db\_xref="GI:7300338"  
/translation="MVKVGSLSPLFEDSPANKINTGDLVNGKKVILFGVPCATP  
GCKTHLPQYSSADELSKQGVDEIVCVSVNDPFVMSANGKEHGAAGKVRLLADPAG  
GTTKALDVTIDLPLGGVRSKRYSLVENGKVTLENPEDGTGLSGSLANNIGKK"  
complement(join(<6163..6507,6567..>6614))  
/gene="CG7215"  
/product="Ct22253"  
/db\_xref="FLYBASE:FBan0007215"  
/db\_xref="FLYBASE:FBgn0038571"  
complement(<6163..>6614)  
/gene="CG7215"  
/map="90E3-90E4"  
/db\_xref="FLYBASE:FBan0007215"  
/db\_xref="FLYBASE:FBgn0038571"  
complement(join(6163..6507,6567..6614))  
/gene="CG7215"  
/note="CG7215 gene product"  
/codon\_start=1  
/db\_xref="FLYBASE:FBan0007215"  
/db\_xref="FLYBASE:FBgn0038571"  
/protein\_id="AAF55498.1"  
/db\_xref="GI:7300339"  
/translation="MOITIKVLKGDCTIEVAPTSTILEVKHIOIEAEOISATNOKLL  
LIGRPLNNQTTASYPNIKECTKLNLVVKPCLRDSILRGFRKHYSLEAERTNEM  
ADPERKINSQSLDRLARLADSIYNRAAT"  
join(7165..7303,9093..9534,10965..>11147)  
/gene="CG7901"  
/product="Ct23882"  
/db\_xref="FLYBASE:FBan0007901"  
/db\_xref="FLYBASE:FBgn0038572"

Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S., Pfeiffer, B., Poon, L., Sequeira, A., Sethi, H., Snir, E., Swirskas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and Rubin, G.M.

# **TITLE**

## **JOURNAL**

## **COMMENT**

Submitted (03-FEB-1999) Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121, Berkeley, CA 94720, USA

On Jan 31, 2000 this sequence version replaced gi:6587780. For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive Web site (<http://www.fruitfly.org/sequence/>) or send email to [bdp@fruitfly.berkeley.edu](mailto:bdp@fruitfly.berkeley.edu). All contigs in this submission meet the following cutoffs: length >= 200 bases.

\* NOTE: This is a 'working draft' sequence. It currently consists of 38 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. \* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 667: contig of 667 bp in length  
 \* 668 747: gap of unknown length  
 \* 748 1288: contig of 541 bp in length  
 \* 1289 1368: gap of unknown length  
 \* 1369 7705: contig of 6337 bp in length  
 \* 7706 7785: gap of unknown length  
 \* 7786 13266: contig of 5481 bp in length  
 \* 13267 13346: gap of unknown length  
 \* 13347 134516: contig of 121170 bp in length  
 \* 134517 134596: gap of unknown length  
 \* 134597 135458: contig of 862 bp in length  
 \* 135459 135538: gap of unknown length  
 \* 135539 135764: contig of 226 bp in length  
 \* 135765 135844: gap of unknown length  
 \* 135845 136080: contig of 236 bp in length  
 \* 136081 136160: gap of unknown length  
 \* 136161 136683: contig of 523 bp in length  
 \* 136684 137663: gap of unknown length  
 \* 137664 137566: contig of 803 bp in length  
 \* 137567 137566: gap of unknown length  
 \* 137647 138152: contig of 506 bp in length  
 \* 138153 138232: gap of unknown length  
 \* 138233 138485: contig of 253 bp in length  
 \* 138486 138565: gap of unknown length  
 \* 138566 139286: contig of 721 bp in length  
 \* 139287 139366: gap of unknown length  
 \* 139367 140013: contig of 647 bp in length  
 \* 140014 140093: gap of unknown length  
 \* 140094 140815: contig of 722 bp in length  
 \* 140816 140895: gap of unknown length  
 \* 140896 141513: contig of 618 bp in length  
 \* 141514 141593: gap of unknown length  
 \* 141594 142372: contig of 779 bp in length  
 \* 142373 142452: gap of unknown length  
 \* 142453 143112: contig of 660 bp in length  
 \* 143113 143192: gap of unknown length  
 \* 143193 144042: contig of 850 bp in length  
 \* 144043 144122: gap of unknown length  
 \* 144123 144683: contig of 561 bp in length  
 \* 144684 14763: gap of unknown length  
 \* 14764 145169: contig of 406 bp in length  
 \* 145170 145249: gap of unknown length  
 \* 145250 146124: contig of 875 bp in length  
 \* 146125 146204: gap of unknown length  
 \* 146205 146979: contig of 775 bp in length  
 \* 146980 147059: gap of unknown length  
 \* 147060 147581: contig of 522 bp in length  
 \* 147582 147661: gap of unknown length  
 \* 147662 148202: contig of 541 bp in length  
 \* 148203 148282: gap of unknown length  
 \* 148283 148505: contig of 223 bp in length  
 \* 148506 148585: gap of unknown length

\* 148586 149025: contig of 440 bp in length  
 \* 149026 149105: gap of unknown length  
 \* 149106 149495: contig of 390 bp in length  
 \* 149496 149575: gap of unknown length  
 \* 149576 150161: contig of 586 bp in length  
 \* 150162 150241: gap of unknown length  
 \* 150242 150736: contig of 495 bp in length  
 \* 150737 150816: gap of unknown length  
 \* 150817 151159: contig of 343 bp in length  
 \* 151160 151239: gap of unknown length  
 \* 151240 152123: contig of 884 bp in length  
 \* 152124 152203: gap of unknown length  
 \* 152204 152586: contig of 383 bp in length  
 \* 152587 152666: gap of unknown length  
 \* 152667 153590: contig of 924 bp in length  
 \* 153591 153670: gap of unknown length  
 \* 153671 154319: contig of 649 bp in length  
 \* 154320 154399: gap of unknown length  
 \* 154400 154743: contig of 344 bp in length  
 \* 154744 154823: gap of unknown length  
 \* 154824 155022: contig of 599 bp in length  
 \* 155023 155502: gap of unknown length  
 \* 155503 155840: contig of 338 bp in length.

## **FEATURES**

### **source**

1. 155840  
 /organism="Drosophila melanogaster"  
 /strain="y: cn bw sp"  
 /db\_xref="taxon:7227"  
 /chromosome="3"  
 /map="90E1-90F10"  
 /clone="BACR48113 (D522) RPCI-98 48.I.13"  
 /clone\_lib="RPCI-98 (Roswell Park Cancer Institute  
 Drosophila melanogaster BAC library, partial EcoRI in  
 pBACE3.6")

BASE COUNT 41735 a 34360 c 34802 g 41426 t 3517 others  
 ORIGIN

### **Alignment Scores:**

Pred. No.: 1.32e+03 Length: 155840  
 Score: 54.00 Matches: 12  
 Percent Similarity: 66.67% Conservative: 4  
 Best Local Similarity: 50.00% Mismatches: 8  
 Query Match: 41.86% Indels: 0  
 DB: 2 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x AC006495 (1-155840)

Oy 5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValIysAla 24  
 ||| |||||::: ||||| :::::|||||::: ||||| |||

Db 13996 GAGGAGGAAGCTGACGAGGAGGAGCGCCCGAGGAGGATGCGCGCTGCCGCC 13937  
 ||| |||||::: ||||| :::::|||||::: ||||| |||

Oy 25 AspAsnAlaAla 28  
 ||| |||||

Db 13936 CATGCTGCCGCG 13925

### **RESULT 14**

AE003721/c

LOCUS AE003721 224896 bp DNA linear INV 05-OCT-2000  
 DEFINITION Drosophila melanogaster genomic scaffold 142000013386035 section 46  
 of 105, complete sequence.

ACCESSION AE003721 AE002708

VERSION AE003721.1 GI:7300335

KEYWORDS HTG.

SOURCE fruit fly.

ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 224896)

Adams, M.D., Celnik, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D.,

Ananides, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., Galle, R.F.,

George, R.A., Lewis, S.E., Richards, S., Ashburner, M., Henderson, S.N.,

Sutton, G.G., Wortman, J.R., Yandell, M.D., Zhang, Q., Chen, L.X.,

For further information on this sequence e-mail to fly@celera.com.  
\* NOTE: This is a 'working draft' sequence.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.

```
FEATURES
  source
    Location/Qualifiers
      /organism="Drosophila melanogaster"
      /db_xref="taxon:7227"
BASE COUNT      9248 a 7110 c 6761 g 9151 t
ORIGIN
Alignment Scores:
  Pred. No.:      259      Length:      32270
  Score:          54.00      Matches:      12
  Percent Similarity: 66.67%      Conservative: 4
  Best Local Similarity: 50.00%      Mismatches: 8
  Query Match:      41.86%      Indels: 0
  DB:              2      Gaps: 0
US-09-847-539A-6_COPY_59_86 (1-28) x AC014470 (1-32270)
QY  5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaValVallyAla 24
||| |||||: ||||| :|||: |||||
Db 26993 GAGGAGGAAGCTGACGAGGAGCGCGCGGAGGATGCCGCGGCTGCCGCC 26934
QY  25 AspAsnAlaAla 28
||| |||||
Db 26933 GATGCTGCGCGC 26922
RESULT 12
AC009462/c
LOCUS
DEFINITION
  Drosophila melanogaster, chromosome 3R, region 90E-90E, BAC clone
  BACR27G04, complete sequence.
AC009462
AC009462.6 GI:15451514
HTG.
SOURCE
  fruit fly.
ORGANISM
  Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
  1 (bases 1 to 151610)
  Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C.,
  Rogers,J.F., An,H., Baldwin,D., Banton,J., Beeson,K.Y., Bissam,D.A.,
  Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M.,
  Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,
  Ferriera,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A.,
  Gonzalez,M., Houch,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
  Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattel,B., Moshrefi,A.,
  McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,
  Pacleb,J., Paragas,V., Parphy,S., Patel,S., Pfeiffer,B.,
  Phouanavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
  Stapleton,M., Strong,R., Svirskaas,R., Tector,C., Williams,S.M.,
  Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
  Sequencing of Drosophila chromosome 3R, region 90E-90E
  Unpublished
  2 (bases 1 to 151610)
  Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
  Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
  Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
  Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
  Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
  Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
  Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
  Svirskaas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
  Rubin,G.M.
  Direct Submission
  Submitted (23-AUG-1999) Drosophila Genome Center, Lawrence Berkeley
  Laboratory, MS 64-121, Berkeley, CA 94720, USA
  On Sep 6, 2001 this sequence version replaced gi:13384332.

FEATURES
  source
    Location/Qualifiers
      /organism="Drosophila melanogaster"
      /db_xref="taxon:7227"
BASE COUNT      43139 a 33567 c 32376 g 42528 t
ORIGIN
Alignment Scores:
  Pred. No.:      1,29e+03      Length:      151610
  Score:          54.00      Matches:      12
  Percent Similarity: 66.67%      Conservative: 4
  Best Local Similarity: 50.00%      Mismatches: 8
  Query Match:      41.86%      Indels: 0
  DB:              3      Gaps: 0
US-09-847-539A-6_COPY_59_86 (1-28) x AC009462 (1-151610)
QY  5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaValVallyAla 24
||| |||||: ||||| :|||: |||||
Db 73153 GAGGAGGAAGCTGACGAGGAGCGCGCGGAGGATGCCGCGGCTGCCGCC 73094
QY  25 AspAsnAlaAla 28
||| |||||
Db 73093 GATGCTGCGCGC 73082
RESULT 13
AC006495/c
LOCUS
DEFINITION
  Drosophila melanogaster chromosome 3 clone BACR48113 (D522) RPCI-98
  48.1.13 map 90E1-90F10 strain Y; cn bw sp, *** SEQUENCING IN
  PROGRESS ***, 38 unordered pieces.
AC006495
AC006495.17 GI:6838793
HTG: HTGS_PHASE1.
SOURCE
  fruit fly.
ORGANISM
  Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
  1 (bases 1 to 155840)
  Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
  Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
  Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
  Hinkle,A., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K.,
  Kearney,L., Lee,B., Lewis,S., Li,P., Ling,H., Moshrefi,A.R.,
  Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B.,
  Richards,S., Sethi,H., Svirskaas,R.R., Wan,K.H., Webster,D.,
  Woolley,P., Yang,S., Yee,M., Yu,C. and Rubin,G.M.
  Sequencing of Drosophila melanogaster
  Unpublished
  2 (bases 1 to 155840)
  Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
  Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
  Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
  Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
  Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
  Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
  Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
  Svirskaas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
  Rubin,G.M.
  Direct Submission
  Submitted (23-AUG-1999) Drosophila Genome Center, Lawrence Berkeley
  Laboratory, MS 64-121, Berkeley, CA 94720, USA
  On Sep 6, 2001 this sequence version replaced gi:13384332.

Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720
This sequence was assembled using end sequences from a whole genome
shotgun and from subclones of this BAC and its neighboring clones.
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive Web site (http://www.fruitfly.org/sequence/) or send email
to bdgpe@fruitfly.berkeley.edu.
Location/Qualifiers
  1..151610
  /organism="Drosophila melanogaster"
  /strain="y; cn bw sp"
  /db_xref="taxon:7227"
  /chromosome="3R"
  /map="90E-90E"
  /clone="BACR27G04 (D985)"
  /clone_lib="RPCI-98 (Roswell Park Cancer Institute
  Drosophila melanogaster BAC library, partial EcoRI ln
  pBACe3.6)"
BASE COUNT      43139 a 33567 c 32376 g 42528 t
ORIGIN
Alignment Scores:
  Pred. No.:      1,29e+03      Length:      151610
  Score:          54.00      Matches:      12
  Percent Similarity: 66.67%      Conservative: 4
  Best Local Similarity: 50.00%      Mismatches: 8
  Query Match:      41.86%      Indels: 0
  DB:              3      Gaps: 0
US-09-847-539A-6_COPY_59_86 (1-28) x AC009462 (1-151610)
QY  5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaValVallyAla 24
||| |||||: ||||| :|||: |||||
Db 73153 GAGGAGGAAGCTGACGAGGAGCGCGCGGAGGATGCCGCGGCTGCCGCC 73094
QY  25 AspAsnAlaAla 28
||| |||||
Db 73093 GATGCTGCGCGC 73082
RESULT 13
AC006495/c
LOCUS
DEFINITION
  Drosophila melanogaster chromosome 3 clone BACR48113 (D522) RPCI-98
  48.1.13 map 90E1-90F10 strain Y; cn bw sp, *** SEQUENCING IN
  PROGRESS ***, 38 unordered pieces.
AC006495
AC006495.17 GI:6838793
HTG: HTGS_PHASE1.
SOURCE
  fruit fly.
ORGANISM
  Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
  1 (bases 1 to 155840)
  Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
  Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
  Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
  Hinkle,A., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K.,
  Kearney,L., Lee,B., Lewis,S., Li,P., Ling,H., Moshrefi,A.R.,
  Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B.,
  Richards,S., Sethi,H., Svirskaas,R.R., Wan,K.H., Webster,D.,
  Woolley,P., Yang,S., Yee,M., Yu,C. and Rubin,G.M.
  Sequencing of Drosophila melanogaster
  Unpublished
  2 (bases 1 to 155840)
  Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
  Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
  Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
  Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
  Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
  Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
  Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
  Svirskaas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
  Rubin,G.M.
  Direct Submission
  Submitted (23-AUG-1999) Drosophila Genome Center, Lawrence Berkeley
  Laboratory, MS 64-121, Berkeley, CA 94720, USA
  On Sep 6, 2001 this sequence version replaced gi:13384332.
```



TITLE	Parry,C., Quail,M., Rutherford,K., Simmonds,M., Skelton,J., Stevens,K., Whitehead,S. and Barrell,B.G.	CDS	complement(375..650) /gene="HCM2.0003c"
JOURNAL	Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhi CT18		/note="HCM2.0003c, hypothetical protein, len: 91 aa; highly similar to TR:CAB52371 (EMBL:AL117211), YPMT1.46HC, Yersinia pestis CO-92 hypothetical protein from plasmid pMT1 (91 aa), fasta scores: E(): 0, 94.5% identity in 91 aa overlap and TR:Q92GX9 (EMBL:AF074611), Y1117, Yersinia pestis KIM5 hypothetical protein from plasmid pMT1 (91 aa), fasta scores: E(): 0, 94.5% identity in 91 aa overlap"
PUBMED	Nature 413 (6858), 848-852 (2001)		/codon_start=1
REFERENCES	11677608		/transl_table=11
AUTHORS	Parkhill,J.	RBS	/product="hypothetical protein"
TITLE	Direct Submission	gene	/protein_id="CAD09870.1"
JOURNAL	Submitted (25-Oct-2001) Submitted on behalf of the Salmonella sequencing team, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk		/db_xref="GI:16505984"
COMMENT	Details of S. typhi sequencing at the Sanger Centre are available on the World Wide Web. (URL, <a href="http://www.sanger.ac.uk/Projects/S_typhi/">http://www.sanger.ac.uk/Projects/S_typhi/</a> ).		/translation="MKILVIRISASTDYVYPLFMVKDCGLNDEETQAAIERNLVEYTGMDADSVYVDDGVCVHNGSCWYVDDTMTVSDEAAHLERILGISTFE"
FEATURES	Location/Qualifiers		complement(658..663)
source	1..106516		/note="possible RBS"
	/organism="Salmonella enterica subsp. enterica serovar Typhi"	gene	complement(719..1129)
	/plasmid="pHCW2"	CDS	/gene="HCM2.0004c"
	/strain="CT18"		complement(719..1129)
	/db_xref="taxon:90370"		/gene="HCM2.0004c"
gene	complement(1..29)		/note="HCM2.0004c, hypothetical protein, len: 136 aa; highly similar to TR:CAB5229 (EMBL:AL117211), YPMT1.47C, Yersinia pestis CO-92 hypothetical protein from plasmid pMT1 (136 aa), fasta scores: E(): 0, 97.8% identity in 136 aa overlap and TR:O68762 (EMBL:AF074611), Y1116, Yersinia pestis KIM5 hypothetical protein from plasmid pMT1 (136 aa), fasta scores: E(): 0, 97.8% identity in 136 aa overlap"
CDS	/gene="HCM2.0001c"		/codon_start=1
	complement(1..29)		/transl_table=11
	/partial		/product="hypothetical protein"
	/gene="HCM2.0001c"		/protein_id="CAD09871.1"
	/note="HCM2.0001c, hypothetical protein, len: 68 aa; highly similar to TR:CAB5227 (EMBL:AL117211), YPMT1.45C, Yersinia pestis CO-92 hypothetical protein from plasmid pMT1 (71 aa), fasta scores: E(): 5.9e-25, 93.3% identity in 60 aa overlap and TR:O68760 (EMBL:AF074611), Y1119, Yersinia pestis KIM5 hypothetical protein from plasmid pMT1 (71 aa), fasta scores: E(): 5.9e-25, 93.3% identity in 60 aa overlap. Spans the end of the sequence"		/db_xref="GI:16505985"
	/codon_start=1		/translation="NKLYDRANAALVDVDTIAOGFPEPEWATOLREALAEMNAPEPSEDEADQORFIRMVAAEEVGPTTAQOAMLLKYPKEAGENLPVDDTPHFWHAWRKFDVIYTRGMSKDMVWHLMHIDRAVDRTLEKFFPPA"
misc_feature	1..232	gene	complement(1113..1484)
	/transl_table=11	CDS	/gene="HCM2.0005c"
	/product="hypothetical protein"		complement(1113..1484)
	/protein_id="CAD09868.1"		/gene="HCM2.0005c"
	/db_xref="GI:16505982"		/note="HCM2.0005c, hypothetical protein, len: 123 aa; similar to TR:Q38665 (EMBL:X78401), ninX, Bacteriophage P22 Nix protein (112 aa), fasta scores: E(): 8.5e-15, 45.6% identity in 114 aa overlap. Highly similar to TR:CAB5230 (EMBL:AL117211), YPMT1.48C, Yersinia pestis CO-92 hypothetical protein from plasmid pMT1 (123 aa), fasta scores: E(): 0, 87.8% identity in 123 aa overlap and TR:Q92GH0 (EMBL:AF074611), Y1115, Yersinia pestis KIM5 hypothetical protein from plasmid pMT1 (123 aa), fasta scores: E(): 0, 87.8% identity in 123 aa overlap"
	/translation="MAVTLAGLE"		/codon_start=1
	complement(29..334)		/transl_table=11
	/gene="HCM2.0002c"		/product="hypothetical protein"
	/note="HCM2.0002c, possible periplasmic protein, len: 101 aa; similar to TR:CAB5228 (EMBL:AL117211), YPMT1.46C, Yersinia pestis hypothetical protein from plasmid pMT1 (111 aa), fasta scores: E(): 2.8e-26, 95.5% identity in 67 aa overlap. The N-terminal 37 aa are highly similar to YPMT1.46AC (EMBL:AL117211) Yersinia pestis hypothetical protein from plasmid pMT1 (59 aa) (86.8% identity in 38 aa overlap) which lies immediately upstream of YPMT1.46C in pMT1. Contains a possible N-terminal signal sequence"		/protein_id="CAD09872.1"
	/codon_start=1		/db_xref="GI:16505986"
	/transl_table=11		/translation="MKNYSEMTDFEINCLVAEATGHRPLISQYKNGKQECGYTAVVAGIPNGAGTFDWCNDPDAWDIYRHRIGVIPARQPGEWRAAHRKVDSSTPQHLIQPNPFRAAMTVLLMOEKKHEETV"
	/product="putative periplasmic protein"	RBS	complement(1137..1141)
	/protein_id="CAD09869.1"		/gene="HCM2.0005c"
	/db_xref="GI:16505983"		/note="possible RBS"
	/translation="MKTVDLTLLLLIVSTAFVLDCIPTGAIRKALAPVNGAMVVDGGEQFYVVRDRLVSSPYLTRKNGKLSGVGDKFYNKSGDYGVAHKNASYLFDCKREV"	RBS	complement(1492..1497)
	complement(36..43)		/note="possible RBS"
	/gene="HCM2.0002c"	repeat_unit	1565..1621
	/note="possible RBS"	gene	/note="repeat 1: 57 bp repeat"
	224..6722		complement(1647..2489)
	/note="96.5% identical to Yersinia pestis CO-92 plasmid pMT1 (EMBL:AL117211) bases 48733..55208"		/gene="HCM2.0006c"
	complement(375..650)	CDS	complement(1647..2489)
	/gene="HCM2.0003c"		/gene="HCM2.0006c"
			/note="HCM2.0006c, possible lipoprotein, len: 280 aa;





/gene="HP1328"  
/note="similar to GB:M26073 SP:P13510 PID:141927 percent identity: 28.85; identified by sequence similarity;  
putative"  
/codon\_start=1  
/transl\_table=11  
/product="cation efflux system protein (czca)"  
/protein\_id="AAD08369.1"  
/db\_xref="GI:2314493"  
/translation="MKRLWALILFSPLEANAQTOEIKKTEAKSOTRFNISTTK  
VIEFSGSRRYALLPENEALIESQLRFGYVEKLYANKTYPIRKGRDLLSYSP  
ELVSAQSELLSFNQOQGAIEKRLKLLGLSENSIEKIISHSKVNQDMTIIYSHFGKI  
IFKPSQNEGSFKQOELFQIDLSQWALVKVNOQDEFLKTHKAILFVEGKIG  
EQETILENIPINKEKMLEARENPNVQIYPNMPAQQEIFQPKMKILPKREAV  
LKIGKATVFKKDDFGLSPLKAIKAVRLSDGSYEILEGLKAGEEVANALFVLDAQON  
NGD"  
7149..10256  
/gene="HP1329"  
/note="HP1329"  
7149..10256  
/gene="HP1329"  
/note="similar to GB:M26073 SP:P13511 PID:141928 percent identity: 31.32; identified by sequence similarity;  
putative"  
/codon\_start=1  
/transl\_table=11  
/product="cation efflux system protein (czca)"  
/protein\_id="AAD08370.1"  
/db\_xref="GI:2314494"  
/translation="MIEKIIDLVSNNKLLTTLVLLIFLASELWAIKSVRLDALPDLSP  
AQVVQITYPNOSKIVQEQVETPLVSTFMSIANIDTVRGISSEGLIYIIFKQGVN  
LYMARLDQLENRVSNPRDAKVEIGSDSTSGWQYALSSDSKNLSDLKVLQDFY  
RYVALGVDGVSEVASGVGVKDEVTLQNDLSLIRYNSLFOVANAINKSNNDTGGGV  
ILENGFEKIIIRSHGYIQSLKDLLEIVYKKEAIPKIKIDIASVRLTPKPRGAANLNG

gene  
CDS

Alignment Scores:

Pred. No.: 30.4 Length: 12496  
Score: 57.00 Matches: 12  
Percent Similarity: 69.57% Conservatives: 4  
Best Local Similarity: 52.17% Mismatches: 7  
Query Match: 44.19% Indels: 0  
DB: 1 Gaps: 0  
US-09-847-539a-6\_COPY\_59\_86 (1-28) x AE000634 (1-12496)  
Qy 4 LeuGluAlaLeuAlaSpGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVallys 23  
Db 5041 ATACAGGCTTCAAGAGCAATTGACGCTTTAAGTTCTCAAGAAAAAGTCGTTAGCAAG 5100  
Qy 24 AlaAspAsn 26  
Db 5101 TGGGATAAC 5109

RESULT 8  
AE001548  
LOCUS  
DEFINITION  
Helicobacter pylori, strain J99 section 109 of 132 of the complete genome. BCT 20-JAN-1999  
ACCESSION  
AE001548 AE001439  
VERSION  
AE001548.1 GI:4155845  
KEYWORDS  
SOURCE  
Helicobacter pylori J99.  
ORGANISM  
Helicobacter pylori J99  
Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;  
Helicobacter.  
REFERENCE  
1 (bases 1 to 12526)  
AUTHORS  
Alm, R.A., Ling, L.S., Moir, D.T., King, B.L., Brown, E.D., Doig, P.C., Smith, D.R., Noonan, B., Guild, B.C., deJonghe, B.L., Carmel, G., Tummino, P.J., Caruso, A., Uria-Nickelsen, M., Mills, D.M., Ives, C., Gibson, R., Merberg, D., Mills, S.D., Jiang, Q., Taylor, D.E., Vovis, G.F., and Trust, T.J.  
TITLE  
Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen Helicobacter pylori  
JOURNAL  
Nature 397 (6715), 176-180 (1999)

99120557  
MEDLINE  
REMARK  
Erratum: [[published erratum appears in Nature 1999 Feb 25:397(6721):719]]  
2 (bases 1 to 12526)  
REFERENCE  
AUTHORS  
King, B.L., Alm, R.A. and Trust, T.J.  
TITLE  
Direct Submission  
JOURNAL  
Submitted (12-JAN-1999) Astra Research Center Boston, 128 Sidney Street, Cambridge, MA 02139, USA  
COMMENT  
Address all correspondence to: hp@arch.us.astro.com or Richard A. Alm, Astra Research Center Boston, 128 Sidney Street, Cambridge, MA, 02139. Lo-See L. Ling, Donald T. Moir, Douglas R. Smith, Braydon C. Guild, Gilles Carmel, Anthony Caruso, Debra M. Mills, Rene Gibson, and Gerald F. Vovis are with Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA, 02453. Qin Jiang and Diane E. Taylor are with the University of Alberta Department of Medical Microbiology and Immunology, Edmonton, Alberta, Canada, T6G 2H7 and the Canadian Bacterial Diseases Network. All other authors are with Astra Research Center Boston, 128 Sidney Street, Cambridge, MA, 02139. Putative identifications, sequence alignments, and name and sequence search capability are available at ARCB's World Wide Web site. (URL: http://www.astro-boston.com/hpylori).  
FEATURES  
source  
1..12526  
/organism="Helicobacter pylori J99"  
/strain="J99"  
/db\_xref="taxon:85963"  
114..1244  
/gene="hjp1241"  
114..1244  
/gene="hjp1241"  
/note="similar to H. pylori 26695 gene HP1321"  
/codon\_start=1  
/transl\_table=11  
/product="putative"  
/protein\_id="AAD06826.1"  
/db\_xref="GI:4155853"  
SDKICLLGPMKSGKTTFAKLAKGPNVYINCNDRLNQNILSKWLEKKMNDL  
LILDRIERDLSPLKPLVILPNTLSHTIPNFSLCVALGNFKEYISFPKPTPKN  
TLFNRFRDGNALDSFTENEQEKILKQENIQALQAFYAPLMARNAKYSLFSAFY  
LYTQLKELKTSKDTLYKLHAKLQRIIFLAPNFENHKTLYLCDFALPYSLTSPS  
LLNVFENMVLELYKQPNVLYAHNCGIFILRNPTSKLALIAHAFTPTPHLEKQLLW  
CHKHGFLLKIAVVSINAPISANNPPYKHLNFIDFSLDIQSILV"  
1509..2087  
/gene="hjp1242"  
1509..2087  
/gene="hjp1242"  
/note="similar to H. pylori 26695 gene HP1322"  
/codon\_start=1  
/transl\_table=11  
/product="putative"  
/protein\_id="AAD06827.1"  
/db\_xref="GI:4155854"  
/translation="MNALKLSFCAALLSLGFAOTVHAQHUKDTINYPDWLKINLFDK  
KNPNQYVGSASISGKNDYFNSYIPYDDKLPPKNAEIALPARMNAIYSLLESALL  
TKMCRNRYKALQVNNVYSHLFGVDFLTSKLSIAKRAEIVDITNHRVYVWVQFPFIQPE  
DLIAYFKAKRIDLSASATNLSALNKALFHL"  
complement(2061..2690)  
/gene="rnhb"  
/note="hjp1243"  
complement(2061..2690)  
/gene="rnhb"  
/note="similar to H. pylori 26695 gene HP1323"  
/codon\_start=1  
/transl\_table=11  
/product="RIBONUCLEASE HII"  
/protein\_id="AAD06828.1"  
/db\_xref="GI:4155855"  
/translation="MGCVSMTLGIDEGAGCGLAGSLFVAGVVCNEKIEFLKMLKMD  
SKLSPKKRFFLEDKIKTHGVEFFVKKSANEIDHGLGACLKLAIIEIVENGCSLA  
NEIKIDGNTAFGLNKRYPNIQTIKGDETIAQIAMASVIAKADREMLIEHALFKEY  
GWDKNCGYGTQKHTEATNKLGAFTPHRHSTLKNRILNPKLLEVEQRLV"

JOURNAL MEDLINE REMARK	Nature 388 (6642), 539-547 (1997) 97394467 Erratum: [[published erratum appears in Nature 1997 Sep 25; 389(6649):412]]
REFERENCE AUTHORS	2 (bases 1 to 12496) Tomb, J.-F., White, O., Kerlavage, A.R., Clayton, R.A., Sutton, G.G., Fleischmann, R.D., Ketchum, K.A., Klenk, H.P., Gill, S., Dougherty, B.A., Nelson, K., Quackenbush, J., Zhou, L., Kirkness, E.F., Peterson, S., Loftus, B., Richardson, D., Dodson, R., Khalak, H.G., Glodek, A., McKenney, K., Fitzgerald, L.M., Lee, N., Adams, M.D., Hickey, E.K., Berg, D.E., Gocayne, J.D., Utterback, T.R., Peterson, J.D., Kelley, J.M., Cotton, M.D., Weidman, J.M., Fujii, C., Bowman, C., Watthey, L., Wallin, E., Hayes, W.S., Borodovsky, M., Karp, P.D., Smith, H.O., Fraser, C.M. and Venter, J.C. Direct Submission Submitted (06-AUG-1997) The Institute for Genomic Research, 9712 Medical Center Dr. Rockville, MD 20850, USA 3 (bases 1 to 12496) White, O.
REFERENCE AUTHORS TITLE	Direct Submission
JOURNAL	Submitted (17-MAR-1999) The Institute for Genomic Research, 9712 Medical Center Dr. Rockville, MD 20850, USA
FEATURES source	Location/Qualifiers 1. .12496 /organism="Helicobacter pylori 26695" /strain="26695" /db_xref="taxon:85962" 113. .1246 /gene="Hpl321" 113. .1246 /gene="Hpl321" 30.81; identified by sequence similarity; putative"
gene	/note="similar to GB:L77117 PID:1592175 percent identity: 30.81; identified by sequence similarity; putative"
CDS	/codon_start=1 /transl_table=11 /product="conserved hypothetical ATP-binding protein" /protein_id="AAD08366.1" /db_xref="GI:2314490" /translation="MRYDYLNPISFERRKFMPLSCLHPFGFPETPKPVLNPLKAPSSDKICLLGPMKSGTKTFALKLAKVKNPVYINNDMLNQNILSSLLKWLKMKMDL LILDRDLRDLSPKPKIVLIPNCLSPITAPNFSICYALGLNFKYTSFPKPTPKN TLNRLRDNALDSLFTENEQEKILKQENIQIIFQAVAPLMKICISQSKFVSAPY LYTQLKRLSKDTLYLKKLHALEKQRIFLFPNENNKTKLYCDFALPYSUTPSPS LNVFNEMVELYKOPPYELYSHDNGIFILRENSTNKIALIAHAFFTPHFLKQLL WCHKHGFLNIIVYSINAPISATNTPYKHLNFIQSLDIQSLV"
gene	1452. .2057 /gene="Hpl322"
CDS	1452. .2057 /gene="Hpl322" /note="hypothetical protein; identified by Genemark; putative" /codon_start=1 /transl_table=11 /product="H. pylori predicted coding region Hpl322" /protein_id="AAD08374.1" /db_xref="GI:2314498"
gene	/translation="MIFYRKATMNALKKLSFCALLSLGLFAQTAHAKHLKGTINPYD WLETFNFEKPNQVYGSASISGRNDYFANTIPYDQLPPEONAEKIALLRARINA YSTLESTLLTKMHNRIYKVLQKNVISHLFLGLVDFLTSKSILAKRFVDVTNHRVYVM VQPFITQPEDLIAYFAKAKRIDLSASATHLSALLNKAFLH"
gene	complement(2031. .2660) /gene="Hpl323"
CDS	complement(2031. .2660) /gene="Hpl323" /note="similar to GB:L42023 SP:P43808 PID:1006316 PID:1221175 PID:1205308 percent identity: 36.32; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="ribonuclease HII (rnhB)" /protein_id="AAD08367.1" /db_xref="GI:2314491"
gene	/translation="MGCVSMVTLGIDEAGRGCLAGSLFVAGVACNEKTALEFLKMLGD
gene	SKKLSLKRRFFLEYIKTKHGEVGFVYVKKSANEDISLGSAGCLAKLAVQRIENGC SLV DEIKIDGNTAFGLNKRYPHIQTIKGDETIAQIAMASVLAKAFKDRMLELHALFREY GNDKNGCYGCKOIHIEAIKLGATPPFHRHSFTLKNRILNPKLLEVEQRLL"
gene	complement(2676. .2927) /gene="Hpl324"
CDS	complement(2676. .2927) /gene="Hpl324" /note="hypothetical protein; identified by Genemark; putative" /codon_start=1 /transl_table=11 /product="H. pylori predicted coding region Hpl324" /protein_id="AAD08375.1" /db_xref="GI:2314499"
gene	/translation="MNPETHCKEIKRILKEALHNDKDLNLYLESKSHAKLTSGAN HLIPSSPSDRKSVKNFEKELTEFIKLRNRRITHTCE"
gene	complement(2949. .4340) /gene="Hpl325"
CDS	complement(2949. .4340) /gene="Hpl325" /note="similar to SP:P05042 GB:X00522 GB:X04065 PID:41513 GB:U00096 percent identity: 63.68; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="fumarase (fumC)" /protein_id="AAD08368.1" /db_xref="GI:2314492"
gene	/translation="MQFRIEHDTMGEIKVNDYWGACOTORSLENEKIGTEKMPKELI GAFKUKRSLAVVNHKLGKLSLEKSAIICACDIKLGELCGEFPFIAIWOTGSGTQTN MNLNEVIANKATEILGNGFRKKLIHPNDDVNMSSNDTFFPAMHIVSVEITHRL PSLNENLKTFKESQOFKEIVTGRTHQADATPLTGLQESFYASHLESHKQIILFSL EHLRELAIGTAVGTGLNHAKESEKVAELQSGVKFVAPNKFHALTSHDAIAYA HGAFKALANLMIANDIRWLASGRCGLGELNIPENPGSSITMPCKVNPTQCEATM VAVQVGNDAIGIAASQGNFELNFKPVIYINFLQSLRLSSDSSEFNHCASGIEP NREKIDYILHLSLMLVLTALNPHVGENAAKIAHNAHKKGISLAKESALELKLJSAEDFD KFYVPEKMIGPKA"
gene	4515. .4892 /gene="Hpl326"
CDS	4515. .4892 /gene="Hpl326" /note="hypothetical protein; identified by Genemark; putative" /codon_start=1 /transl_table=11 /product="H. pylori predicted coding region Hpl326" /protein_id="AAD08376.1" /db_xref="GI:2314500"
gene	/translation="MKRLAALFLVSLGVNGLNANEOTLKANDLEVKIKSVGNPIKGD NTFILSPTLKGKALEIAIVRFQMPMPGMPAMKEMAOVSEKNGLYEAKTNLSMNGT WQVRVDITKSGEQVYRAKTSIDL"
gene	4897. .6135 /gene="Hpl327"
CDS	4897. .6135 /gene="Hpl327" /note="hypothetical protein; identified by Genemark; putative" /codon_start=1 /transl_table=11 /product="H. pylori predicted coding region Hpl327" /protein_id="AAD08377.1" /db_xref="GI:2314501"
gene	/translation="MLSFISAFDKRGVSIRLLTALLLLSLGLAKDLQISQVAKYLS KNQIQALQEQIDALSQEKVSKWDPNIPILGYNNANVSDFRFLDSTLMQNMISGLS KQVDLNGKLTQSQMDLEKQKILELKTQOLATSLMINGIENYKNOQEIQLKTA IKNLEWTLYQANSHSSPNLJIAKLEILKQLEIKKNLEALSAHSYSGELAFREN ELLSIAPKNFENFNRQELHNSATINVDIAFARDEKSKDITLAKRSFLIEDNVNGLV YFYSKOYNNYMFESIALSTPLPIYQKALVQEKKESLSVFSKSEVFNKNKTHHAL KLLKLETLQKLNESINKIKQEKIAQIYALDLKNSGYNAYNFAFNDRITITQITOL ETLNLSNLTSLSLQNLKGL"
gene	6132. .7148 /gene="Hpl328"
CDS	6132. .7148



```
source      1. .832
/organism="Streptococcus pyogenes"
/strain="ATCC700294"
/db_xref="ATCC:700294"
/db_xref="taxon:1314"
1. .99
1. .654
/feature="grab"
/codon_start=1
/transl_table=11
/product="GRAB precursor"
/protein_id="AAD26338.1"
/db_xref="GI:4589079"
/transl_table="MGKIKVKCFRRSAFLGVAASVLGVSTVSAVDSP IEQPRII
PNGTLLNLGNAPKIALNEERAIDELKQAIEDKEATTATIEAASDALEALADOT
DALQSEAAVVKADNAADALEALADOTDALQSEAEVVDSDNAASDAWEKAATPIAL
DVKKTKTPVVKKEERQNVNLTPTGSESNPFTAAALAMVSTGVLVVSSCKEN"
100. .651
mat_peptide 304 a 143 c 169 g 216 t
100. .273
misc_feature 274. .441
/feature="Region: repeat motif"
442. .576
misc_feature 577. .651
/feature="Region: cell-wall attachment"
577. .651
misc_feature 304 a 143 c 169 g 216 t
/feature="Region: membrane-spanning"
BASE COUNT 304 a 143 c 169 g 216 t
ORIGIN
1.41e-12      Length:      832
Score:      129.00      Matches:      28
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match: 100.00%      Indels: 0
Db: 1      Gaps: 0

Alignment Scores:
Pred. No.:      1.41e-12      Length:      832
Score:      129.00      Matches:      28
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match: 100.00%      Indels: 0
Db: 1      Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x AF124399 (1-832)

Qy      1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluAlaAla 20
Db      274 TCAGATGCGCTTAGAGCATTACCGGATCAACAGACGCTTTACAATCAGAGAAGCTGCG 333
Qy      21 ValValLysAlaAspAsnAlaAla 28
Db      334 GTTGTAAAGCGGATAACGCTGCT 357

RESULT 6
AE006573/c
LOCUS
DEFINITION Streptococcus pyogenes M1 GAS strain SF370, section 102 of 167 of
the complete genome.
ACCESSION AE006573 AE004092
VERSION AE006573.1 GI:13622459
KEYWORDS
SOURCE
ORGANISM Streptococcus pyogenes M1 GAS.
Bacteria: Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
Streptococcus.
1 (bases 1 to 10029)
Ferretti,J.J., McShan,W.M., Adjic,D., Savic,D., Savic,G., Lyon,K.,
Primeaux,C., Sezate,S.S., Surorov,A.N., Kenton,S., Lai,H., Lin,S.,
Qian,Y., Jia,H.G., Najjar,F.Z., Ren,Q., Zhu,H., Song,L., White,J.,
Yuan,X., Clifton,S.W., Roe,B.A. and McLaughlin,R.E.
Complete genome sequence of an M1 strain of Streptococcus pyogenes
Proc. Natl. Acad. Sci. U.S.A. 98 (8), 4658-4663 (2001)
21192684
MEDLINE 11296296
PUBMED
REFERENCE 2 (bases 1 to 10029)
Ferretti,J.J., McShan,W.M., Adjic,D., Savic,D., Savic,G., Lyon,K.,
Primeaux,C., Sezate,S.S., Surorov,A.N., Kenton,S., Lai,H., Lin,S.,
Qian,Y., Jia,H.G., Najjar,F.Z., Ren,Q., Zhu,H., Song,L., White,J.,
Yuan,X., Clifton,S.W., Roe,B.A. and McLaughlin,R.E.
Complete genome sequence of an M1 strain of Streptococcus pyogenes
Proc. Natl. Acad. Sci. U.S.A. 98 (8), 4658-4663 (2001)

TITLE Direct Submission
JOURNAL Submitted (10-APR-2001) Department of Microbiology and Immunology,
University of Oklahoma Health Sciences Center, 940 SL Young Blvd.,
Oklahoma City, OK 73104, USA
FEATURES
source
1. .10029
/organism="Streptococcus pyogenes M1 GAS"
/strain="Sp370"
/serotype="M1"
/db_xref="taxon:160490"
complement(135. .626)
/gene="arok"
/feature="SPyl351"
complement(135. .626)
/gene="arok"
/EC_number="2.7.1.71"
/feature="Best Blastp hit = pir|D82693 shikimate kinase
xfl335 [imported] - xylella fastidiosa (strain 9a5c)
>gil9106332|gb|AAF84144.1|AE003966_5 (AE003966) shikimate
kinase [Xylella fastidiosa]"
/codon_start=1
/transl_table=11
/product="putative shikimate kinase"
/protein_id="AAK34179.1"
/db_xref="GI:13622460"
/transl_table="MTKVLGFMGVGKTTVSKHLMCKMDAIIIEAKIGMSIAAFIE
QHGEIAFRTIESOVKLDLIFANDNSIIVTGGVVVLOENRLLKHNHHTLLVASPE
TLYQRLAKHDKSRPLFLKYSKEAFYEFYQORVMVYEGSLDLIVRDHRTPEEVANII
EGY"
complement(619. .1911)
/gene="aroA"
/feature="SPyl352"
complement(619. .1911)
/gene="aroA"
/EC_number="2.5.1.19"
/feature="Best Blastp hit = sp|P43905|AROA_LACLA
3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE
(5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE) (EPSP
SYNTHASE) (EPSPS) >gil1075724|pir||S52580
3-phosphoshikimate 1-carboxyvinyltransferase (EC 2.5.1.19)
- Lactococcus lactis >gil683583|emb|CAA55180.1| (X78413)
5-enolpyruvylshikimate-3-phosphate synthase [Lactococcus
lactis]"
/codon_start=1
/transl_table=11
/product="putative 3-phosphoshikimate
1-carboxyvinyltransferase"
/protein_id="AAK34180.1"
/db_xref="GI:13622461"
/transl_table="MKRMKLRNAGPLOGTIOVPGDKSISHRAVILGAVAKGETRVKG
LLGEDVLSTIQAFNGLVRIEKKDQDVIEGQFQGLNAPCOTLNMGNSCTSMRLIA
GILAQPFVSKMIGDESLSKRPMDRIYPLQMGVEISGETDRQFPPLQLOGNRLQIP
ITYLPISSAOVKSAILLAQAKQTOVKEITRNHTEEMIQFGGLRLLVDGRKIT
LVGPQQLTAQETIYVPGDISSAFLVAGLIIPGSELLKKNVAPRTGILLEVKMG
AQIVYEDMKKEQVTSIRVYVSNMKGITISGGLIPRLIDELPFIALLATQAQGTCTIK
DAQELRVKETDRIQVTDILNSMGANIKATADGMIIKGPVTLYGANTSTYGDHRTGM
TAIALLVKQGVHLDKDEADMTSVPTFFKDLERLCHD"
complement(2013. .2978)
/gene="SPyl353"
complement(2013. .2978)
/feature="Best Blastp hit = dbj|BAB07074.1| (AP001518)
unknown conserved protein [Bacillus halodurans]"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK34181.1"
/db_xref="GI:13622462"
/transl_table="MAEKKWFDKVLVSKWVEPIQVFMRHLSAEMDLISAIVAYVLIIL
TAFPLIVIANIIFPVLNIDIALRLMKQPKDIERPASAIVENIESKPSGSLGVA
TATGLWNTSRSLTSLOKAINKAYGASQRDFHGLVGLTSLIILFLLALPALSIF
SKAAIQVLDKRDHYHLSDNITITIFLLLIQIPITVLIIFVGLMLLFLPLPNVKIKRRIILP
```

REFERENCE 2 (bases 1 to 717)  
AUTHORS Rasmussen,M., Muller,H.P. and Bjorck,L.  
TITLE Direct Submission  
JOURNAL Submitted (28-JAN-1999) Cell and Molecular Biology, Molecular Pathogenesis, Soelvegatan 39, Lund 221 00, Sweden

FEATURES  
source Location/Qualifiers  
1..717  
/organism="Streptococcus pyogenes"  
/strain="KT19"  
/db\_xref="taxon:1314"  
1..>717  
/product="GRAB"  
1..174  
/note="Region: alpha2-macroglobulin-binding A domain"  
CDS  
1..>717  
/note="grab"  
/codon\_start=1  
/transl\_table=11  
/product="GRAB precursor"  
/protein\_id="AAD26341.1"  
/db\_xref="GI:4589085"  
/translation="VDSPIEQPRIIPNGTTLNLLGNAPEKALRNEERAIDELKKQA  
IEDKEATTAIEAASSDALEADQDALQSEEAHVQSDNAASDALEALADQDALQSE  
EAAVKADNAASDTLEALADQDALQSEEAHVQSDNAASDTLEALADQDALQSE  
AAVKADNAASDTLEALADQDALQSEEAHVQSDNAASDTLEALADQDALQSE  
TRPVKKERQNVNLTPTTGE"  
misc\_feature 175..594  
/note="Region: repeat motif"  
misc\_feature 595..717  
/note="Region: cell-wall attachment"  
BASE COUNT 257 a 140 c 163 g 157 t  
ORIGIN

Alignment Scores:  
Pred. No.: 1,21e-12 Length: 717  
Score: 129.00 Matches: 28  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 1 Gaps: 0

US-09-847-539a-6\_COPY\_59\_86 (1-28) x AF124402 (1-717)  
Qy 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAla 20  
|||||  
Db 259 AGTGACGCCCTTAGAGCATTTGGCGATCAACACAGACGCTTTACAATCAGAAGAGCTGCG 318  
Qy 21 ValVallyAlaAspAsnAlaAla 28  
|||||  
Db 319 GTTGTTAAAGCGGATACGCTGCT 342

RESULT 4  
AF124403  
LOCUS AF124403 804 bp DNA linear BCT 14-AUG-2000  
DEFINITION Streptococcus pyogenes strain AP49 GRAB precursor, gene, partial cds.  
ACCESSION AF124403  
VERSION AF124403.1 GI:4589086  
KEYWORDS  
SOURCE Streptococcus pyogenes.  
ORGANISM Streptococcus pyogenes  
Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
Streptococcus.  
1 (bases 1 to 804)  
Rasmussen,M., Muller,H.P. and Bjorck,L.  
TITLE Protein GRAB of streptococcus pyogenes regulates proteolysis at the bacterial surface by binding alpha2-macroglobulin  
J. Biol. Chem. 274 (22), 15336-15344 (1999)  
JOURNAL MEDLINE  
PUBMED 99269061  
10336419  
REFERENCE 2 (bases 1 to 804)  
AUTHORS Rasmussen,M., Muller,H.P. and Bjorck,L.  
TITLE Direct Submission  
JOURNAL Submitted (28-JAN-1999) Cell and Molecular Biology, Molecular Pathogenesis, Soelvegatan 39, Lund 221 00, Sweden

JOURNAL Submitted (28-JAN-1999) Cell and Molecular Biology, Molecular Pathogenesis, Soelvegatan 39, Lund 221 00, Sweden

FEATURES  
source Location/Qualifiers  
1..804  
/organism="Streptococcus pyogenes"  
/strain="AP49"  
/db\_xref="taxon:1314"  
1..>804  
/product="GRAB"  
1..174  
/note="Region: alpha2-macroglobulin-binding A domain"  
CDS  
1..>804  
/note="grab"  
/codon\_start=1  
/transl\_table=11  
/product="GRAB precursor"  
/protein\_id="AAD26342.1"  
/db\_xref="GI:4589087"  
/translation="VDSPIEQPRIIPNGTTLNLLGNAPEKALRNEERAIDELKKQA  
IEDKEATTAIEAASSDALEADQDALQSEEAHVQSDNAASDALEALADQDALQSE  
EAAVKADNAASDTLEALADQDALQSEEAHVQSDNAASDTLEALADQDALQSE  
AAVKADNAASDTLEALADQDALQSEEAHVQSDNAASDTLEALADQDALQSE  
VQSDNAASDAMEKAATPIALDVKTKDTPVVKKEERQNVNLTPTTGE"  
misc\_feature 175..678  
/note="Region: repeat motif"  
misc\_feature 679..804  
/note="Region: cell-wall attachment"  
BASE COUNT 284 a 161 c 185 g 174 t  
ORIGIN

Alignment Scores:  
Pred. No.: 1,36e-12 Length: 804  
Score: 129.00 Matches: 28  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 1 Gaps: 0

US-09-847-539a-6\_COPY\_59\_86 (1-28) x AF124403 (1-804)  
Qy 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAla 20  
|||||  
Db 511 AGTGACGCCCTTAGAGCATTTGGCGATCAACACAGACGCTTTACAATCAGAAGAGCTGCG 570  
Qy 21 ValVallyAlaAspAsnAlaAla 28  
|||||  
Db 571 GTTGTTAAAGCGGATACGCTGCT 594

RESULT 5  
AF124399  
LOCUS AF124399 832 bp DNA linear BCT 14-AUG-2000  
DEFINITION Streptococcus pyogenes strain ATCC700294 GRAB precursor, gene, complete cds.  
ACCESSION AF124399  
VERSION AF124399.1 GI:4589078  
KEYWORDS  
SOURCE Streptococcus pyogenes.  
ORGANISM Streptococcus pyogenes  
Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
Streptococcus.  
1 (bases 1 to 832)  
Rasmussen,M., Muller,H.P. and Bjorck,L.  
TITLE Protein GRAB of streptococcus pyogenes regulates proteolysis at the bacterial surface by binding alpha2-macroglobulin  
J. Biol. Chem. 274 (22), 15336-15344 (1999)  
JOURNAL MEDLINE  
PUBMED 99269061  
10336419  
REFERENCE 2 (bases 1 to 832)  
AUTHORS Rasmussen,M., Muller,H.P. and Bjorck,L.  
TITLE Direct Submission  
JOURNAL Submitted (28-JAN-1999) Cell and Molecular Biology, Molecular Pathogenesis, Soelvegatan 39, Lund 221 00, Sweden

```

ORGANISM      Streptococcus pyogenes
               Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
               Streptococcus.
REFERENCE     1 (bases 1 to 468)
AUTHORS      Rasmussen,M., Muller,H.P. and Bjorck,L.
TITLE        Protein GRAB of streptococcus pyogenes regulates proteolysis at the
               bacterial surface by binding alpha2-macroglobulin
JOURNAL      J. Biol. Chem. 274 (22), 15336-15344 (1999)
MEDLINE      99269061
PUBMED       10336419
REFERENCE     2 (bases 1 to 468)
AUTHORS      Rasmussen,M., Muller,H.P. and Bjorck,L.
TITLE        Direct Submission
JOURNAL      Submitted (28-JAN-1999) Cell and Molecular Biology, Molecular
               Pathogenesis, Soelvegatan 39, Lund 221 00, Sweden
FEATURES     Location/Qualifiers
               source
               1..468
               /organism="Streptococcus pyogenes"
               /strain="KTL3"
               /db_xref="taxon:1314"
               mat_peptide
               1..>468
               /product="GRAB"
               misc_feature
               1..174
               /note="Region: alpha2-macroglobulin-binding A domain"
               CDS
               <1..>468
               /note="grab"
               /codon_start=1
               /transl_table=11
               /product="GRAB precursor"
               /protein_id="AAD26340.1"
               /db_xref="GI:4589083"
               /translation="VDSPIEQPRIIPNGTILNLGNAPKALRNEAIDELKKQA
               IEDKATTAIEAASDALEALADQTDALQSEEAAYVKADNAASDALEALADQTDALOS
               EEAEVQSDNAASDAWEKAATPIALDVKTKDTKPVVKEERQNVNTLPTTGE"
               misc_feature
               175..342
               /note="Region: repeat motif"
               misc_feature
               343..468
               /note="Region: cell-wall attachment"
               BASE COUNT 178 a 91 c 100 g 99 t
               ORIGIN
               Alignment Scores:
               Pred. No.: 7 76e-13 Length: 468
               Score: 129.00 Matches: 28
               Percent Similarity: 100.00% Conservative: 0
               Best Local Similarity: 100.00% Mismatches: 0
               Query Match: 100.00% Indels: 0
               DB: 1 Gaps: 0
               US-09-847-539A-6_COPY_59_86 (1-28) x AFL24401 (1-468)
               QY 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAla 20
               Db 175 TCAGATGCCTTAGAAGCATTAGCGGATCAACAGACGCTTTACAATCAGAGAAGCTGCG 234
               QY 21 ValValIysAlaAspAsnAlaAla 28
               Db 235 GTTGTAAAGCGGATAACGCTGCT 258
               RESULT 2
               AF124400
               LOCUS AF124400
               DEFINITION Streptococcus pyogenes strain AP1 GRAB precursor, gene, partial
               cds.
               ACCESSION AF124400
               VERSION AF124400.1 GI:4589080
               KEYWORDS
               SOURCE Streptococcus pyogenes.
               ORGANISM Streptococcus pyogenes
               Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
               Streptococcus.
               1 (bases 1 to 469)
               Rasmussen,M., Muller,H.P. and Bjorck,L.
               TITLE Protein GRAB of streptococcus pyogenes regulates proteolysis at the
               bacterial surface by binding alpha2-macroglobulin
               JOURNAL J. Biol. Chem. 274 (22), 15336-15344 (1999)
               MEDLINE 99269061
               PUBMED 10336419

```

```

TITLE Protein GRAB of streptococcus pyogenes regulates proteolysis at the
JOURNAL J. Biol. Chem. 274 (22), 15336-15344 (1999)
MEDLINE 99269061
PUBMED 10336419
REFERENCE 2 (bases 1 to 469)
AUTHORS Rasmussen,M., Muller,H.P. and Bjorck,L.
TITLE Direct Submission
JOURNAL Submitted (28-JAN-1999) Cell and Molecular Biology, Molecular
               Pathogenesis, Soelvegatan 39, Lund 221 00, Sweden
FEATURES Location/Qualifiers
               source
               1..469
               /organism="Streptococcus pyogenes"
               /strain="API"
               /db_xref="taxon:1314"
               mat_peptide
               1..>469
               /product="GRAB"
               misc_feature
               1..174
               /note="Region: alpha2-macroglobulin-binding A domain"
               CDS
               <1..>469
               /note="grab"
               /codon_start=1
               /transl_table=11
               /product="GRAB precursor"
               /protein_id="AAD26339.1"
               /db_xref="GI:4589081"
               /translation="VDSPIEQPRIIPNGTILNLGNAPKALRNEAIDELKKQA
               IEDKATTAIEAASDALEALADQTDALQSEEAAYVKADNAASDALEALADQTDALOS
               EEAEVQSDNAASDAWEKAATPIALDVKTKDTKPVVKEERQNVNTLPTTGE"
               misc_feature
               175..342
               /note="Region: repeat motif"
               misc_feature
               343..469
               /note="Region: cell-wall attachment"
               BASE COUNT 178 a 90 c 100 g 101 t
               ORIGIN
               Alignment Scores:
               Pred. No.: 7 77e-13 Length: 469
               Score: 129.00 Matches: 28
               Percent Similarity: 100.00% Conservative: 0
               Best Local Similarity: 100.00% Mismatches: 0
               Query Match: 100.00% Indels: 0
               DB: 1 Gaps: 0
               US-09-847-539A-6_COPY_59_86 (1-28) x AFL24400 (1-469)
               QY 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAla 20
               Db 175 TCAGATGCCTTAGAAGCATTAGCGGATCAACAGACGCTTTACAATCAGAGAAGCTGCG 234
               QY 21 ValValIysAlaAspAsnAlaAla 28
               Db 235 GTTGTAAAGCGGATAACGCTGCT 258
               RESULT 3
               AF124402
               LOCUS AF124402
               DEFINITION Streptococcus pyogenes strain KTL9 GRAB precursor, gene, partial
               cds.
               ACCESSION AF124402
               VERSION AF124402.1 GI:4589084
               KEYWORDS
               SOURCE Streptococcus pyogenes.
               ORGANISM Streptococcus pyogenes
               Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
               Streptococcus.
               1 (bases 1 to 717)
               Rasmussen,M., Muller,H.P. and Bjorck,L.
               TITLE Protein GRAB of streptococcus pyogenes regulates proteolysis at the
               bacterial surface by binding alpha2-macroglobulin
               JOURNAL J. Biol. Chem. 274 (22), 15336-15344 (1999)
               MEDLINE 99269061
               PUBMED 10336419

```

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 13, 2002, 02:12:06 ; Search time 273.262 Seconds  
(without alignments)  
2144.253 Million cell updates/sec

Title: US-09-847-539A-6\_COPY\_59\_86

Perfect score: 129

Sequence: 1 SDALEALADQTDALQSEEAHVVKADNAA 28

Scoring table:

BLOSUM62  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool/US09847539/runat\_10102002\_092548\_3423/app\_query.fasta\_1.526  
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=40 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09847539.ecgn\_1\_1182.erunat\_10102002\_092548\_3423 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

GenEmbl.\*

- 1: gb\_ba.\*
- 2: gb\_htg.\*
- 3: gb\_in.\*
- 4: gb\_om.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vi.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_mu.\*
- 20: em\_om.\*
- 21: em\_or.\*
- 22: em\_ov.\*
- 23: em\_pat.\*
- 24: em\_ph.\*
- 25: em\_pl.\*
- 26: em\_ro.\*
- 27: em\_sts.\*
- 28: em\_un.\*

- 29: em\_vi.\*
- 30: em\_htg\_hum.\*
- 31: em\_htg\_inv.\*
- 32: em\_htg\_other.\*
- 33: em\_htgo\_inv.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	129	100.0	468	1	AF124401 Streptoco
2	129	100.0	469	1	AF124400 Streptoco
3	129	100.0	717	1	AF124402 Streptoco
4	129	100.0	804	1	AF124403 Streptoco
5	129	100.0	832	1	AF124399 Streptoco
6	129	100.0	10029	1	AE006573 Streptoco
7	57	44.2	12496	1	AE000634 Helicobac
8	57	44.2	12526	1	AE001548 Helicobac
9	57	44.2	106516	1	STYPPHCM2 AF031321 Tetrahyme
10	54	41.9	1288	3	AF031321 Tetrahyme
11	54	41.9	32270	2	AC014470 Drosophil
12	54	41.9	151610	3	AC009462 Drosophil
13	54	41.9	155840	2	AC006495 Drosophil
14	54	41.9	224896	3	AE003721 Drosophil
15	53	41.1	7686	1	SCOA14671 Streptomy
16	53	41.1	31317	1	SCSCP2 AL645771 Streptomy
17	53	41.1	117309	9	AC007274 Homo sapi
18	53	41.1	173772	2	AC087862 Homo sapi
19	52	40.3	12373	3	DME17920 Drosophila
20	52	40.3	16135	3	AF096897 Drosophil
21	52	40.3	17595	3	AF096896 Drosophil
22	52	40.3	36993	2	AC018007 Drosophil
23	52	40.3	133155	2	AC011910 Drosophil
24	52	40.3	169427	3	AC007416 Drosophil
25	52	40.3	281993	3	AE003619 Drosophil
26	52	40.3	303100	1	RME603646 Rhizobium
27	51	39.5	2151	3	AY061146 Drosophil
28	51	39.5	3838	3	PPAS7 M10130 Plasmodium
29	51	39.5	12423	3	AE005155 Halobacte
30	51	39.5	14941	3	AF072878 Tetrahyme
31	51	39.5	20217	1	AE000047 Mycoplasma
32	51	39.5	30561	1	SC1F2 AL031350 Streptomy
33	51	39.5	96288	3	AC005749 Drosophil
34	51	39.5	105823	2	AC019871 Drosophil
35	51	39.5	135509	8	AC037426 Oryza sat
36	51	39.5	175115	3	AC105899 Drosophil
37	51	39.5	183898	3	AC092244 Drosophil
38	51	39.5	200050	1	AL591973 Listeria
39	51	39.5	202050	1	AL646065 Ralstonia
40	51	39.5	231450	1	AL596163 Listeria
41	51	39.5	324484	3	AE003584 Drosophil
42	50	38.8	1298	1	ECODCDA M90069 E.coli deox
43	50	38.8	1409	1	ECU11035 U1035 Escherichia
44	50	38.8	10224	1	AE005432 Escherich
45	50	38.8	10780	3	CEMYO2 X08065 Caenorhabd

ALIGNMENTS

RESULT 1  
AF124401  
LOCUS AF124401 468 bp DNA linear BCT 14-AUG-2000  
DEFINITION Streptococcus pyogenes strain KTL3 GRAB precursor, gene, partial cds.  
ACCESSION AF124401  
VERSION AF124401.1 GI:4589082  
KEYWORDS Streptococcus pyogenes.  
SOURCE



**THIS PAGE BLANK (USPTO)**

```

; APPLICATION NUMBER: <Unknown>
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 191:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1033 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 191:
US-08-961-083-191

```

```

Alignment Scores:
Pred. No.:      97.9      Length:      1033
Score:          44.00     Matches:      11
Percent Similarity: 65.00%   Conservative: 2
Best Local Similarity: 55.00% Mismatches: 7
Query Match:      34.11%    Indels:      0
DB:               4.      Gaps:        0

```

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-08-961-083-191 (1-1033)

```

QY      8 AlaAspGlnThrAspAlaLeuGlnSerGluAlaAlaValValLysAlaAspAsnAla 27
||| ||||| ||||| ||| ||| ||||| ||| ||| ||||| ||| |||
Db      8 GCAGGGCAGACAGATGCCTCGCAATTGAAAGGGCGCAGTTAGCCCAAGGAGGAAAAGCA 67

```

Search completed: October 13, 2002, 04:45:47  
Job time : 99.2834 secs

; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 172754  
; LENGTH: 862  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-172754

Alignment Scores:  
Pred. No.: 79.4 Length: 862  
Score: 44.00 Matches: 9  
Percent Similarity: 65.22% Conservative: 6  
Best Local Similarity: 39.13% Mismatches: 8  
Query Match: 34.11% Indels: 0  
DB: 7 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-10-027-632-172754 (1-862)

Qy 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAala 20  
Db 46 ACACAGACACTCCAGAGTCTCAGCAGTCAGCTACTTCTCTGCAGGYAGAAAGCAGAGA 105  
Qy 21 ValVallys 23  
Db 106 GTTCTAAAG 114

RESULT 38

US-10-027-632-172755  
; Sequence 172755, Application US/10027632

; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; FILE REFERENCE: 108827,129

; CURRENT APPLICATION NUMBER: US/10/027,632  
; PRIOR FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 172755  
; LENGTH: 862  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-172755

Alignment Scores:  
Pred. No.: 79.4 Length: 862  
Score: 44.00 Matches: 9  
Percent Similarity: 65.22% Conservative: 6  
Best Local Similarity: 39.13% Mismatches: 8  
Query Match: 34.11% Indels: 0  
DB: 7 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-10-027-632-172755 (1-862)  
Qy 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAala 20  
Db 46 ACACAGACACTCCAGAGTCTCAGCAGTCAGCTACTTCTCTGCAGGYAGAAAGCAGAGA 105  
Qy 21 ValVallys 23  
Db 106 GTTCTAAAG 114

RESULT 39

US-10-155-881-5807  
; Sequence 5807, Application US/10155881  
; GENERAL INFORMATION:  
; APPLICANT: Dolson, Stanton B.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Liu, Jingdong  
; APPLICANT: Luflyya, Linda L.  
; APPLICANT: McIninch, James  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH  
; TITLE OF INVENTION: TRANSCRIPTION IN PLANTS  
; FILE REFERENCE: 38-21(15300)J  
; CURRENT APPLICATION NUMBER: US/10/155,881  
; CURRENT FILING DATE: 2002-05-22  
; NUMBER OF SEQ ID NOS: 37595  
; SEQ ID NO 5807  
; LENGTH: 940  
; TYPE: DNA  
; ORGANISM: Zea mays  
US-10-155-881-5807

Alignment Scores:

Pred. No.: 87.8 Length: 940  
Score: 44.00 Matches: 9  
Percent Similarity: 67.86% Conservative: 10  
Best Local Similarity: 32.14% Mismatches: 9  
Query Match: 34.11% Indels: 0  
DB: 7 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-10-155-881-5807 (1-940)

Qy 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAala 20  
Db 335 CGGAGGCGCTGGCGGGTCTCTCCAAAGTACGAGCGCTGCGCAGCCAGCACACGAG 394  
Qy 21 ValVallysAlaAspAlaAala 28  
Db 395 ATGTGGCGCGCTGCTCGCGCTCG 418

RESULT 40

US-08-961-083-191  
; Sequence 191, Application US/08961083

; GENERAL INFORMATION:  
; APPLICANT: Choi et. al.  
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines

; NUMBER OF SEQUENCES: 452  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
; COMPUTER: HP Vectra 486/33  
; OPERATING SYSTEM: MSDOS version 6.2  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/961,083  
; FILING DATE: 30-Oct-1997  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-09-918-995-23409 (1-546)

QY 1 SerAspAlaLeuGlnThrAspAlaLeuGlnSerGluGluAlaA 20  
DB 159 GCAGAGCTCTGCCCTTGGCGGAGGAGCCATCTTGTGCAAAACACGAAACCAAG 218  
QY 21 ValVallysAlaAspAsnAlaA 28  
DB 219 GTGGCTATGGCGGTTCCAGCAGCA 242

## RESULT 34

PCT-US02-27884-4854/c  
; Sequence 4854, Application PC/TUS0227884  
; GENERAL INFORMATION:  
; APPLICANT: The Dow Chemical Company  
; TITLE OF INVENTION: Nucleic Acid Compositions Conferring Altered Metabolic Character  
; FILE REFERENCE: DOM-07611  
; CURRENT APPLICATION NUMBER: PCT/US02/27884  
; CURRENT FILING DATE: 2002-08-30  
; NUMBER OF SEQ ID NOS: 7560  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 4854  
; LENGTH: 656  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Synthetic  
PCT-US02-27884-4854

Alignment Scores:  
Pred. No.: 57.9 Length: 656  
Score: 44.00 Matches: 11  
Percent Similarity: 60.00% Conservative: 1  
Best Local Similarity: 55.00% Mismatches: 8  
Query Match: 34.11% Indels: 0  
DB: 1 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x PCT-US02-27884-4854 (1-656)

QY 5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaValVallysAla 24  
DB 120 GAGCGCGATGCCGACGCCCGCGCCGACGACAGATGCGGCGCGGTGAATGCC 61

## RESULT 35

PCT-US02-27884-5555/c  
; Sequence 5555, Application PC/TUS0227884  
; GENERAL INFORMATION:  
; APPLICANT: The Dow Chemical Company  
; TITLE OF INVENTION: Nucleic Acid Compositions Conferring Altered Metabolic Character  
; FILE REFERENCE: DOM-07611  
; CURRENT APPLICATION NUMBER: PCT/US02/27884  
; CURRENT FILING DATE: 2002-08-30  
; NUMBER OF SEQ ID NOS: 7560  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 5555  
; LENGTH: 656  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Synthetic  
PCT-US02-27884-5555

Alignment Scores:  
Pred. No.: 57.9 Length: 656  
Score: 44.00 Matches: 11  
Percent Similarity: 60.00% Conservative: 1  
Best Local Similarity: 55.00% Mismatches: 8  
Query Match: 34.11% Indels: 0  
DB: 1 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x PCT-US02-27884-5555 (1-656)

QY 5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaValVallysAla 24  
DB 120 GAGCGCGATGCCGACGCCCGCGCCGACGACGAAGAAGATGAGCGCGCGTGAATGCC 61

## RESULT 36

US-10-027-632-172753  
; Sequence 172753, Application US/10027632  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 172753  
; LENGTH: 862  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-172753

Alignment Scores:  
Pred. No.: 79.4 Length: 862  
Score: 44.00 Matches: 9  
Percent Similarity: 65.22% Conservative: 6  
Best Local Similarity: 39.13% Mismatches: 8  
Query Match: 34.11% Indels: 0  
DB: 7 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-10-027-632-172753 (1-862)

QY 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaA 20  
DB 46 ACACAGACACTCCAGAGTCTCAGACGTCAGACTCATTTCTGCGAGGAGAAAGCAGAGA 105

QY 21 ValVallys 23

DB 106 GTTCTAAG 114

## RESULT 37

US-10-027-632-172754  
; Sequence 172754, Application US/10027632  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218

```
; NUMBER OF SEQ ID NOS: 350
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 137
; LENGTH: 425
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (398)
; OTHER INFORMATION: n equals a,t,g, or c
US-10-217-489-137

Alignment Scores:
Pred. No.: 35 Length: 425
Score: 44.00 Matches: 10
Percent Similarity: 60.87% Conservative: 4
Best Local Similarity: 43.48% Mismatches: 9
Query Match: 34.11% Indels: 0
DB: 6 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x US-10-217-489-137 (1-425)
Oy 5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValLysAla 24
Db 101 CAGGCATTAGAGCCCAACACAGGATGCTTCCAGGCCCAAGACGACGATGATGGAT 160
Oy 25 AspAsnAla 27
Db 161 CAGAAGGCA 169

RESULT 31
US-10-221-279-4262/c
; Sequence 4262, Application US/10221279
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-046
; CURRENT APPLICATION NUMBER: US/10/221,279
; CURRENT FILING DATE: 2002-09-06
; PRIOR APPLICATION NUMBER: 09/574,454
; PRIOR FILING DATE: 2000-05-19
; PRIOR APPLICATION NUMBER: 09/519,705
; PRIOR FILING DATE: 2000-03-07
; NUMBER OF SEQ ID NOS: 12360
; SOFTWARE: Custom
; SEQ ID NO 4262
; LENGTH: 489
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-221-279-4262

Alignment Scores:
Pred. No.: 41.2 Length: 489
Score: 44.00 Matches: 10
Percent Similarity: 61.34% Conservative: 6
Best Local Similarity: 38.46% Mismatches: 10
Query Match: 34.11% Indels: 0
DB: 6 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x US-10-221-279-4262 (1-489)
Oy 3 AlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaVal 22
Db 114 GCGACAGAGCTCTGGTGGAGCCACTAGTCCTCCAGGACGACGAGGCGGAGATCCTG 55
Oy 23 LysAlaAspAsnAlaAla 28
Db 54 ACCGCTGACGGCGACGA 37

RESULT 32
US-10-027-632-13710/c
; Sequence 13710, Application US/10027632
; GENERAL INFORMATION:
```

```
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13710
; LENGTH: 511
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-13710

Alignment Scores:
Pred. No.: 43.3 Length: 511
Score: 44.00 Matches: 8
Percent Similarity: 81.25% Conservative: 5
Best Local Similarity: 50.00% Mismatches: 3
Query Match: 34.11% Indels: 0
DB: 7 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x US-10-027-632-13710 (1-511)
Oy 2 AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGlu 17
Db 132 GAGGCCATGGAAGCAACGCCTGATCAAACTGACTTAACTGCAGAC 85

RESULT 33
US-09-918-995-23409
; Sequence 23409, Application US/09918995
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 23409
; LENGTH: 546
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(546)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-23409

Alignment Scores:
Pred. No.: 46.8 Length: 546
Score: 44.00 Matches: 10
Percent Similarity: 60.71% Conservative: 7
Best Local Similarity: 35.71% Mismatches: 11
Query Match: 34.11% Indels: 0
DB: 5 Gaps: 0
```

Percent Similarity: 63.64% Conservative: 3  
Best Local Similarity: 50.00% Mismatches: 8  
Query Match: 34.89% Indels: 0  
DB: 8 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-60-360-039-44171 (1-3435)

QY 3 AlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVal 22  
DB 1345 GCCCAGGAGCTGCTGCCGAGCAGGAAGCGGCTGTCAGGAGCGGAGCTGCACAAATC 1404

QY 23 LysAla 24

DB 1405 GCCGCG 1410

RESULT 27

US-09-975-254-11911  
; Sequence 11911, Application US/09975254  
; GENERAL INFORMATION:  
; APPLICANT: Byrum, Joseph R.  
; APPLICANT: Heck, Gregory R.  
; APPLICANT: La Rosa, Thomas J.  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with  
; FILE OF INVENTION: Plants  
; FILE REFERENCE: 38-21(15309)B  
; CURRENT APPLICATION NUMBER: US/09/975,254  
; CURRENT FILING DATE: 2001-10-12  
; PRIOR APPLICATION NUMBER: US/09/263,191  
; PRIOR FILING DATE: 1999-03-05  
; NUMBER OF SEQ ID NOS: 31255  
; SEQ ID NO 11911  
; LENGTH: 237  
; TYPE: DNA  
; ORGANISM: Glycine max  
; OTHER INFORMATION: Clone ID: 700905432H1  
US-09-975-254-11911

Alignment Scores:  
Pred. No.: 17.8 Length: 237  
Score: 44.00 Matches: 9  
Percent Similarity: 52.38% Conservative: 2  
Best Local Similarity: 42.86% Mismatches: 10  
Query Match: 34.11% Indels: 0  
DB: 5 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-09-975-254-11911 (1-237)

QY 5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLysAla 24

DB 153 GAGACCTTCCTTCACAAACTCTCNCCTTGAGAGCAACCTNCCATACNNAAGC 212

QY 25 Asp 25

DB 213 GAC 215

RESULT 28

US-09-918-995-37583/c  
; Sequence 37583, Application US/09918995  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc.  
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
; FILE OF INVENTION: FROM VARIOUS CDNA LIBRARIES  
; FILE REFERENCE: 20411-756  
; CURRENT APPLICATION NUMBER: US/09/918,995  
; CURRENT FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US/09/235,076  
; PRIOR FILING DATE: 1999-01-20  
; NUMBER OF SEQ ID NOS: 38054  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 37583  
; LENGTH: 370  
; TYPE: DNA  
; ORGANISM: Homo sapiens

US-09-918-995-37583

Alignment Scores:  
Pred. No.: 29.8 Length: 370  
Score: 44.00 Matches: 10  
Percent Similarity: 60.87% Conservative: 4  
Best Local Similarity: 43.48% Mismatches: 9  
Query Match: 34.11% Indels: 0  
DB: 5 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-09-918-995-37583 (1-370)

QY 5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLysAla 24

DB 129 CAGGCATTAGAGCCCAACACAGGATGCTTTCCAGGCCAAGACGACGACGTAATGATGGAT 70

QY 25 AspAsnAla 27

DB 69 CAGAAGGCA 61

RESULT 29

US-09-918-995-37603/c  
; Sequence 37603, Application US/09918995  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc.  
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
; FILE OF INVENTION: FROM VARIOUS CDNA LIBRARIES  
; FILE REFERENCE: 20411-756  
; CURRENT APPLICATION NUMBER: US/09/918,995  
; CURRENT FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US/09/235,076  
; PRIOR FILING DATE: 1999-01-20  
; NUMBER OF SEQ ID NOS: 38054  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 37603  
; LENGTH: 373  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-918-995-37603

Alignment Scores:  
Pred. No.: 30.1 Length: 373  
Score: 44.00 Matches: 10  
Percent Similarity: 60.87% Conservative: 4  
Best Local Similarity: 43.48% Mismatches: 9  
Query Match: 34.11% Indels: 0  
DB: 5 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-09-918-995-37603 (1-373)

QY 5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLysAla 24

DB 129 CAGGCATTAGAGCCCAACACAGGATGCTTTCCAGGCCAAGACGACGACGTAATGATGGAT 70

QY 25 AspAsnAla 27

DB 69 CAGAAGGCA 61

RESULT 30

US-10-217-489-137  
; Sequence 137, Application US/10217489  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE OF INVENTION: PM045CIN  
; CURRENT APPLICATION NUMBER: US/10/217,489  
; CURRENT FILING DATE: 2002-08-14  
; PRIOR APPLICATION NUMBER: 09/757,030  
; PRIOR FILING DATE: 2001-01-09  
; PRIOR APPLICATION NUMBER: 60/179,065  
; PRIOR FILING DATE: 2000-01-31  
; PRIOR APPLICATION NUMBER: 60/180,628  
; PRIOR FILING DATE: 2000-02-04



Score: 45.00 Matches: 11  
Percent Similarity: 61.54% Conservative: 5  
Best Local Similarity: 42.31% Mismatches: 10  
Query Match: 34.88% Indels: 0  
DB: 7 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-10-027-632-164207 (1-825)

QY 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluAlaAla 20  
Db 180 GCAGATGCTTTGGAGATGCTGCTTCACAGACTGAAGCCTGTTGTCAAACAGGCTGTCA 121  
QY 21 ValValLysAlaAspAsn 26  
Db 120 CCTATCAGAGTTTAAAT 103

#### RESULT 20

US-10-027-632-164208/c  
; Sequence 164208, Application US/10027632

; GENERAL INFORMATION:

; APPLICANT: Wang, David G.

; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide

; FILE REFERENCE: Polymorphisms in the Human Genome

; CURRENT APPLICATION NUMBER: US/10/027,632

; PRIOR FILING DATE: 2002-04-30

; PRIOR APPLICATION NUMBER: US 60/218,006

; PRIOR FILING DATE: 2000-07-12

; PRIOR APPLICATION NUMBER: US 60/198,676

; PRIOR FILING DATE: 2000-04-20

; PRIOR APPLICATION NUMBER: US 60/193,483

; PRIOR FILING DATE: 2000-03-29

; PRIOR APPLICATION NUMBER: US 60/185,218

; PRIOR FILING DATE: 2000-02-24

; PRIOR APPLICATION NUMBER: US 60/167,363

; PRIOR FILING DATE: 1999-11-23

; PRIOR APPLICATION NUMBER: US 60/156,358

; PRIOR FILING DATE: 1999-09-28

; PRIOR APPLICATION NUMBER: US 60/146,002

; PRIOR FILING DATE: 1999-08-09

; NUMBER OF SEQ ID NOS: 325720

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 164208

; LENGTH: 825

; TYPE: DNA

; ORGANISM: Human

US-10-027-632-164208

#### Alignment Scores:

Pred. No.: 48.9 Length: 825  
Score: 45.00 Matches: 11  
Percent Similarity: 61.54% Conservative: 5  
Best Local Similarity: 42.31% Mismatches: 10  
Query Match: 34.88% Indels: 0  
DB: 7 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-10-027-632-164208 (1-825)

QY 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluAlaAla 20  
Db 180 GCAGATGCTTTGGAGATGCTGCTTCACAGACTGAAGCCTGTTGTCAAACAGGCTGTCA 121  
QY 21 ValValLysAlaAspAsn 26  
Db 120 CCTATCAGAGTTTAAAT 103

#### RESULT 21

US-10-219-999-22677

; Sequence 22677, Application US/10219999

; GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei

; APPLICANT: Edgerton, Michael D

; APPLICANT: Hinkle, Gregory J.

; APPLICANT: Kovalic, David K.  
; APPLICANT: Liu, Jingdong  
; APPLICANT: Stein, Joshua  
; TITLE OF INVENTION: CDNA SEQUENCES AND USES FOR PLANT IMPROVEMENT  
; FILE REFERENCE: 38-10(52726)C  
; CURRENT APPLICATION NUMBER: US/10/219,999  
; CURRENT FILING DATE: 2002-08-15  
; PRIOR APPLICATION NUMBER: US 60/324,109  
; PRIOR FILING DATE: 2001-09-21  
; PRIOR APPLICATION NUMBER: US 60/312,544  
; PRIOR FILING DATE: 2001-08-15  
; NUMBER OF SEQ ID NOS: 63520  
; SEQ ID NO 22677  
; LENGTH: 950  
; TYPE: DNA  
; ORGANISM: Zea mays  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (200)..(538)  
; OTHER INFORMATION:  
US-10-219-999-22677

#### Alignment Scores:

Pred. No.: 57.6 Length: 950  
Score: 45.00 Matches: 10  
Percent Similarity: 63.64% Conservative: 4  
Best Local Similarity: 45.45% Mismatches: 8  
Query Match: 34.88% Indels: 0  
DB: 6 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-10-219-999-22677 (1-950)

QY 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluAlaAla 20  
Db 188 TCGTCGGCGGTGGAGACGGCGCGGTGGAGCGGCTGCGGTGCGAGAGCGCGGTG 247  
QY 21 ValVal 22  
Db 248 GTGGTG 253

#### RESULT 22

US-60-360-039-27473/c

; Sequence 27473, Application US/60360039

; GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei

; APPLICANT: Chen, Xianfeng

; APPLICANT: Goldman, Barry S.

; APPLICANT: Hinkle, Gregory J.

; APPLICANT: Slater, Steven C.

; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

; FILE REFERENCE: 38-10(52052)A

; CURRENT APPLICATION NUMBER: US/60/360,039

; CURRENT FILING DATE: 2002-02-21

; NUMBER OF SEQ ID NOS: 47374

; SEQ ID NO 27473

; LENGTH: 1557

; TYPE: DNA

; ORGANISM: Neurospora crassa

US-60-360-039-27473

#### Alignment Scores:

Pred. No.: 102 Length: 1557  
Score: 45.00 Matches: 9  
Percent Similarity: 66.67% Conservative: 7  
Best Local Similarity: 37.50% Mismatches: 8  
Query Match: 34.88% Indels: 0  
DB: 8 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-60-360-039-27473 (1-1557)

QY 2 AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluAlaAlaVal 21  
Db 188 TCGTCGGCGGTGGAGACGGCGCGGTGGAGCGGCTGCGGTGCGAGAGCGCGGTG 247





US-09-847-539A-6\_COPY\_59\_86 (1-28) x PCT-US02-25766-4301 (1-128978)

QY 8 AlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLysAla 24

Db 63278 GCTGAACAACTGACTCTTAGAACCACTGAGGCACAGGTCTTCAAAGCG 63328

## RESULT 13

US-10-161-453-6

; Sequence 6, Application US/10161453

; GENERAL INFORMATION:

; APPLICANT: EXELIXIS, INC.

; TITLE OF INVENTION: LINKS AS MODIFIERS OF THE p53 PATHWAY AND METHODS OF USE

; FILE REFERENCE: EX02-084C

; CURRENT APPLICATION NUMBER: US/10/161,453

; CURRENT FILING DATE: 2002-06-03

; PRIOR APPLICATION NUMBER: US 60/296,076

; PRIOR FILING DATE: 2001-06-05

; PRIOR APPLICATION NUMBER: US 60/328,605

; PRIOR FILING DATE: 2001-10-10

; PRIOR APPLICATION NUMBER: US 60/357,253

; PRIOR FILING DATE: 2002-02-15

; NUMBER OF SEQ ID NOS: 25

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 6

; LENGTH: 128978

; TYPE: DNA

; ORGANISM: Homo sapiens

US-10-161-453-6

## Alignment Scores:

Pred. No.:	7.16e+03	Length:	128978
Score:	47.00	Matches:	10
Percent Similarity:	76.47%	Conservative:	3
Best Local Similarity:	58.82%	Mismatches:	4
Query Match:	36.43%	Indels:	0
DB:	7	Gaps:	0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-10-161-453-6 (1-128978)

QY 8 AlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLysAla 24

Db 63278 GCTGAACAACTGACTCTTAGAACCACTGAGGCACAGGTCTTCAAAGCG 63328

## RESULT 14

US-10-155-881-3217/c

; Sequence 3217, Application US/10155881

; GENERAL INFORMATION:

; APPLICANT: Dotson, Stanton B.

; APPLICANT: Kovalic, David K.

; APPLICANT: Liu, Jingdong

; APPLICANT: Lufliyya, Linda L.

; APPLICANT: McIninch, James

; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH

; FILE REFERENCE: 38-21(15300)J

; CURRENT APPLICATION NUMBER: US/10/155,881

; CURRENT FILING DATE: 2002-05-22

; NUMBER OF SEQ ID NOS: 37595

; SEQ ID NO 3217

; LENGTH: 705

; TYPE: DNA

; ORGANISM: Zea mays

; FEATURE:

; NAME/KEY: unsure

; LOCATION: {1}..(705)

; OTHER INFORMATION: unsure at all n locations

US-10-155-881-3217

## Alignment Scores:

Pred. No.:	26.4	Length:	705
Score:	46.00	Matches:	10
Percent Similarity:	69.23%	Conservative:	8

Best Local Similarity:	38.46%	Mismatches:	8
Query Match:	35.66%	Indels:	0
DB:	7	Gaps:	0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-10-155-881-3217 (1-705)

QY 3 AlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVal 22

Db 476 GCTGTCGTCGCTGTTGGCCGCGCAAGCGAGGCAGTTCGTGAGGATGATGCCGCTGGCGTC 417

QY 23 LysAlaAspAsnAlaAla 28

Db 416 GAGGCTGACAACGGCGCC 399

## RESULT 15

US-10-219-999-6030/c

; Sequence 6030, Application US/10219999

; GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei

; APPLICANT: Edgerton, Michael D

; APPLICANT: Hinkle, Gregory J.

; APPLICANT: Kovalic, David K.

; APPLICANT: Liu, Jingdong

; APPLICANT: Stein, Joshua

; TITLE OF INVENTION: CDNA SEQUENCES AND USES FOR PLANT IMPROVEMENT

; FILE REFERENCE: 38-10(52726)C

; CURRENT APPLICATION NUMBER: US/10/219,999

; CURRENT FILING DATE: 2002-08-15

; PRIOR APPLICATION NUMBER: US 60/324,109

; PRIOR FILING DATE: 2001-09-21

; PRIOR APPLICATION NUMBER: US 60/312,544

; PRIOR FILING DATE: 2001-08-15

; NUMBER OF SEQ ID NOS: 63520

; SEQ ID NO 6030

; LENGTH: 1512

; TYPE: DNA

; ORGANISM: Zea mays

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (1)..(1410)

; OTHER INFORMATION:

US-10-219-999-6030

## Alignment Scores:

Pred. No.:	64	Length:	1512
Score:	46.00	Matches:	10
Percent Similarity:	69.23%	Conservative:	8
Best Local Similarity:	38.46%	Mismatches:	8
Query Match:	35.66%	Indels:	0
DB:	6	Gaps:	0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-10-219-999-6030 (1-1512)

QY 3 AlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVal 22

Db 1314 GCTGTCGTCGCTGTTGGCCGCGCAAGCGAGGCAGTTCGTGAGGATGATGCCGCTGGCGTC 1255

QY 23 LysAlaAspAsnAlaAla 28

Db 1254 GAGGCTGACAACGGCGCC 1237

## RESULT 16

US-60-360-039-44539

; Sequence 44539, Application US/60360039

; GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei

; APPLICANT: Chen, Xianfeng

; APPLICANT: Goldman, Barry S.

; APPLICANT: Hinkle, Gregory J.

; APPLICANT: Slater, Steven C.

; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

; FILE REFERENCE: 38-10(52052)A

QY 22 VallysAlaAspAsn 26  
:::  
Db 2235 CTGGCGAGCATAAC 2221

## RESULT 9

US-09-440-302B-636

; Sequence 636, Application US/09440302B  
; GENERAL INFORMATION:  
; APPLICANT: Chenchik, Alex  
; TITLE OF INVENTION: Human Neurobiology Array  
; FILE REFERENCE: CLON-006CIP11  
; CURRENT APPLICATION NUMBER: US/09/440.302B  
; PRIOR FILING DATE: 1999-11-17  
; PRIOR APPLICATION NUMBER: 09/053,375  
; PRIOR FILING DATE: 1998-03-31  
; NUMBER OF SEQ ID NOS: 1193  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 636  
; LENGTH: 6965  
; TYPE: DNA  
; ORGANISM: Homo Sapiens  
US-09-440-302B-636

Alignment Scores:  
Pred. No.: 243 Length: 6965  
Score: 47.00 Matches: 11  
Percent Similarity: 70.00% Conservative: 3  
Best Local Similarity: 55.00% Mismatches: 6  
Query Match: 36.43% Indels: 0  
DB: 5 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-09-440-302B-636 (1-6965)

QY 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVallys 23

Db 1460 TTGGAAAGCCTTAAGCATCAGCAGGATGCCCTTTGGCTGAAGAAGCTCCAAAGCTTTAAAG 1519  
|||||:||||| ||| ||||||||| :||||| |||:|||||

## RESULT 10

US-10-135-696-3/c

; Sequence 3, Application US/10135696  
; GENERAL INFORMATION:  
; APPLICANT: YAN, Chunhua et al.  
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC  
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES  
; FILE REFERENCE: THEREOF  
; FILE REFERENCE: CL001178DIV  
; CURRENT APPLICATION NUMBER: US/10/135,696  
; CURRENT FILING DATE: 2002-05-01  
; PRIOR APPLICATION NUMBER: 09/813,817  
; PRIOR FILING DATE: 2001-03-22  
; PRIOR APPLICATION NUMBER: 09/978,197  
; PRIOR FILING DATE: 2001-10-17  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 59065  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-135-696-3

Alignment Scores:  
Pred. No.: 2,9e+03 Length: 59065  
Score: 47.00 Matches: 10  
Percent Similarity: 76.47% Conservative: 3  
Best Local Similarity: 58.82% Mismatches: 4  
Query Match: 36.43% Indels: 0  
DB: 7 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-10-135-696-3 (1-59065)

QY 8 AlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVallysAla 24  
|||||:|||||:|||||:||||| ||| |||||||

Db 21239 GCTGAACAACTGACTCCTTAGAACACAGCTGAGGCACAGGTCTTCAAAGCG 21189

## RESULT 11

PCT-US01-17423-6

; Sequence 6, Application PC/TUS0117423  
; GENERAL INFORMATION:  
; APPLICANT: EXELIXIS, INC.  
; TITLE OF INVENTION: LINKS AS MODIFIERS OF THE p53 PATHWAY AND METHODS OF USE  
; FILE REFERENCE: EX02-084C-PC  
; CURRENT APPLICATION NUMBER: PCT/US01/17423  
; CURRENT FILING DATE: 2002-06-03  
; PRIOR APPLICATION NUMBER: US 60/296,076  
; PRIOR FILING DATE: 2001-06-05  
; PRIOR APPLICATION NUMBER: US 60/328,605  
; PRIOR FILING DATE: 2001-10-10  
; PRIOR APPLICATION NUMBER: US 60/357,253  
; PRIOR FILING DATE: 2002-02-15  
; NUMBER OF SEQ ID NOS: 25  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 6  
; LENGTH: 128978  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
PCT-US01-17423-6

Alignment Scores:  
Pred. No.: 7,16e+03 Length: 128978  
Score: 47.00 Matches: 10  
Percent Similarity: 76.47% Conservative: 3  
Best Local Similarity: 58.82% Mismatches: 4  
Query Match: 36.43% Indels: 0  
DB: 1 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x PCT-US01-17423-6 (1-128978)

QY 8 AlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVallysAla 24  
|||||:|||||:|||||:||||| ||| |||||||

Db 63278 GCTGAACAACTGACTCCTTAGAACACAGCTGAGGCACAGGTCTTCAAAGCG 63328

## RESULT 12

PCT-US02-25766-4301

; Sequence 4301, Application PC/TUS0225766  
; GENERAL INFORMATION:  
; APPLICANT: GENE LOGIC, INC.  
; APPLICANT: MUNGER, William E  
; APPLICANT: FAULK, Ronald  
; APPLICANT: SUN, Hongwei  
; APPLICANT: SASAI, Hitoshi  
; APPLICANT: WAGA, Iwao  
; APPLICANT: YAMAMOTO, Jun  
; TITLE OF INVENTION: Gene Expression Profiles in Glomerular Diseases  
; FILE REFERENCE: 44921-5068-WO  
; CURRENT APPLICATION NUMBER: PCT/US02/25766  
; CURRENT FILING DATE: 2002-08-14  
; PRIOR APPLICATION NUMBER: US 60/311,837  
; PRIOR FILING DATE: 2001-08-14  
; NUMBER OF SEQ ID NOS: 13946  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 4301  
; LENGTH: 128978  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: Genbank Accession No. AC002073  
PCT-US02-25766-4301

Alignment Scores:  
Pred. No.: 7,16e+03 Length: 128978  
Score: 47.00 Matches: 10  
Percent Similarity: 76.47% Conservative: 3  
Best Local Similarity: 58.82% Mismatches: 4  
Query Match: 36.43% Indels: 0  
DB: 1 Gaps: 0

```
; APPLICANT: Slater, Steven C.
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)A
; CURRENT APPLICATION NUMBER: US/60/360,039
; CURRENT FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 40312
; LENGTH: 1449
; TYPE: DNA
; ORGANISM: Bacillus thuringiensis
US-60-360-039-40312
Alignment Scores:
Pred. No.: 25 6 Length: 1449
Score: 48.00 Matches: 10
Percent Similarity: 65.22% Conservative: 5
Best Local Similarity: 43.48% Mismatches: 8
Query Match: 37.21% Indels: 0
DB: 8 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x US-60-360-039-40312 (1-1449)
Qy 2 AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaVal 21
:: ||| ||| ||| ||| ::|||::|||::||| ||| |||
Db 1225 GAAGTACTTCGTCGATTGCTGGTCAATAGAACCTCTAGAACCAATGGAGCCGACGCA 1284
Qy 22 Vallysala 24
::||| |||
Db 1285 ATTAAGGCA 1293

RESULT 6
US-10-179-131-3446
; GENERAL INFORMATION:
; APPLICANT: HARE, ROBERTA S.
; APPLICANT: SHAW, KAREN J.
; APPLICANT: SHIMER JR., GEORGE H.
; APPLICANT: KESSLER, MARCO
; APPLICANT: NOLLING, JORK
; APPLICANT: ZENG, QIANDONG
; APPLICANT: GREENE, JONATHAN R.
; TITLE OF INVENTION: CANDIDA ALBICANS NUCLEIC ACIDS AND POLYPEPTIDES,
; FILE REFERENCE: 2976-4031
; CURRENT APPLICATION NUMBER: US/10/179,131
; CURRENT FILING DATE: 2002-06-21
; NUMBER OF SEQ ID NOS: 10194
; SEQ ID NO 3446
; LENGTH: 6093
; TYPE: DNA
; ORGANISM: Candida albicans
US-10-179-131-3446
Alignment Scores:
Pred. No.: 135 Length: 6093
Score: 48.00 Matches: 13
Percent Similarity: 70.83% Conservative: 4
Best Local Similarity: 54.17% Mismatches: 6
Query Match: 37.21% Indels: 1
DB: 7 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x US-10-179-131-3446 (1-6093)
Qy 2 AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAla-Va 21
||| ::|||::|||::||| ||| ||| ||| ||| ::
Db 2125 GATATATCAAGGCTTAGCAGATCCAGTTGACTCTCTGTTATCAACCTCAGCTGCCAAT 2184
Qy 21 lVallysala 24
:|||||
Db 2185 TGTGAAGCG 2194

RESULT 7
```

```
US-60-360-039-40814/C
; Sequence 40814, Application US/60360039
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Chen, Xianfeng
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)A
; CURRENT APPLICATION NUMBER: US/60/360,039
; CURRENT FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 40814
; LENGTH: 783
; TYPE: DNA
; ORGANISM: Bacillus halodurans
US-60-360-039-40814
Alignment Scores:
Pred. No.: 19.4 Length: 783
Score: 47.00 Matches: 10
Percent Similarity: 63.64% Conservative: 4
Best Local Similarity: 45.45% Mismatches: 8
Query Match: 36.43% Indels: 0
DB: 8 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x US-60-360-039-40814 (1-783)
Qy 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVallys 23
||| ||| ||| ||| ::|||::|||::||| ||| ||| ||| ||| ::|||
Db 691 CTGCGTCCTTTGGCTCACCAGACGACCGAATGGAACCGAGGAGCAATCTTGTTC 632
Qy 24 AlaAsp 25
|||
Db 631 ACTGAT 626

RESULT 8
US-60-360-039-27525/c
; Sequence 27525, Application US/60360039
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Chen, Xianfeng
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)A
; CURRENT APPLICATION NUMBER: US/60/360,039
; CURRENT FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 27525
; LENGTH: 4885
; TYPE: DNA
; ORGANISM: Neurospora crassa
US-60-360-039-27525
Alignment Scores:
Pred. No.: 161 Length: 4885
Score: 47.00 Matches: 10
Percent Similarity: 60.00% Conservative: 5
Best Local Similarity: 40.00% Mismatches: 10
Query Match: 36.43% Indels: 0
DB: 8 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x US-60-360-039-27525 (1-4885)
Qy 2 AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaVal 21
||| ||| ||| ||| ::|||::|||::||| ||| ||| ||| ||| ::|||
Db 2295 GACGGCCTCGAACGAGTAGCGAGACTACATAGTACCAACCGACGACAGATCTT 2236
```

```
Query Match: 38.76% Indels: 0
DB: 6 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x US-10-219-999-12394 (1-1679)

Qy 3 AlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaVal 22
||||| :|||:|||| :|||:||||| :|||:||||| |||
Db 1255 GCTGTCGCTGCTGTTGCCGCGAAGCGAGCGAGTGTCTGAGGATGATGCCGCTGGCGTC 1196

Qy 23 LysAlaAspAsnAlaAla 28
:::||||| |||
Db 1195 GAGGCTGACACGGCGCC 1178

RESULT 2
US-10-219-999-8111/c
; Sequence 8111, Application US/10219999
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Edgerton, Michael D
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jingdong
; APPLICANT: Stein, Joshua
; TITLE OF INVENTION: cDNA SEQUENCES AND USES FOR PLANT IMPROVEMENT
; FILE REFERENCE: 38-10(52726)C
; CURRENT APPLICATION NUMBER: US/10/219,999
; PRIOR FILING DATE: 2002-08-15
; PRIOR APPLICATION NUMBER: US 60/324,109
; PRIOR FILING DATE: 2001-09-21
; PRIOR APPLICATION NUMBER: US 60/312,544
; PRIOR FILING DATE: 2001-08-15
; NUMBER OF SEQ ID NOS: 63520
; SEQ ID NO 8111
; LENGTH: 1696
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (3)..(1427)
; OTHER INFORMATION:
US-10-219-999-8111

Alignment Scores:
Pred. No.: 12.9 Length: 1696
Score: 50.00 Matches: 11
Percent Similarity: 73.08% Conservative: 8
Best Local Similarity: 42.31% Mismatches: 7
Query Match: 38.76% Indels: 0
DB: 6 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x US-10-219-999-8111 (1-1696)

Qy 3 AlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaVal 22
||||| :|||:|||| :|||:||||| :|||:||||| |||
Db 1331 GCTGTCGCTGCTGTTGCCGCGAAGCGAGCGAGTGTCTGAGGATGATGCCGCTGGCGTC 1272

Qy 23 LysAlaAspAsnAlaAla 28
:::||||| |||
Db 1271 GAGGCTGACACGGCGCC 1254

RESULT 3
US-10-155-881-3337/c
; Sequence 3337, Application US/10155881
; GENERAL INFORMATION:
; APPLICANT: Dotson, Stanton B.
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jingdong
; APPLICANT: Lutfiyya, Linda L.
; APPLICANT: McIninch, James
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; TITLE OF INVENTION: TRANSCRIPTION IN PLANTS
; FILE REFERENCE: 38-21(15300)J
; CURRENT APPLICATION NUMBER: US/10/155,881
```

```
; CURRENT FILING DATE: 2002-05-22
; NUMBER OF SEQ ID NOS: 37595
; SEQ ID NO 3337
; LENGTH: 1783
; TYPE: DNA
; ORGANISM: Zea mays
US-10-155-881-3337

Alignment Scores:
Pred. No.: 13.7 Length: 1783
Score: 50.00 Matches: 11
Percent Similarity: 73.08% Conservative: 8
Best Local Similarity: 42.31% Mismatches: 7
Query Match: 38.76% Indels: 0
DB: 7 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x US-10-155-881-3337 (1-1783)

Qy 3 AlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaVal 22
||||| :|||:|||| :|||:||||| :|||:||||| |||
Db 1354 GCTGTCGCTGCTGTTGCCGCGAAGCGAGCGAGTGTCTGAGGATGATGCCGCTGGCGTC 1295

Qy 23 LysAlaAspAsnAlaAla 28
:::||||| |||
Db 1294 GAGGCTGACACGGCGCC 1277

RESULT 4
US-60-360-039-30630
; Sequence 30630, Application US/60360039
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Chen, Xianfeng
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)A
; CURRENT APPLICATION NUMBER: US/60/360,039
; CURRENT FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 30630
; LENGTH: 3257
; TYPE: DNA
; ORGANISM: Caenorhabditis elegans
US-60-360-039-30630

Alignment Scores:
Pred. No.: 27.5 Length: 3257
Score: 50.00 Matches: 9
Percent Similarity: 69.57% Conservative: 7
Best Local Similarity: 39.13% Mismatches: 7
Query Match: 38.76% Indels: 0
DB: 8 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x US-60-360-039-30630 (1-3257)

Qy 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAla 20
||||| :|||:|||| :|||:||||| :|||:||||| |||
Db 1915 TCAGATGCTATTCAAGAGCTTTCGATCAATCGAGCAACTTCAAAAGCAAAAGGAAGA 1974

Qy 21 ValValLys 23
::: |||
Db 1975 ATTGAGAAG 1983

RESULT 5
US-60-360-039-40312
; Sequence 40312, Application US/60360039
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Chen, Xianfeng
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
```

GenCore version 5.1.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 13, 2002, 02:41:57 ; Search time 65.2834 Seconds  
(without alignments)  
1453.830 Million cell updates/sec

Title: US-09-847-539a-6\_COPY\_59\_86

Perfect score: 129

Sequence: 1 SDALALADOTDALQSEAAVVKADNAA 28

Scoring table:

BLOSUM62  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2385415 seqs, 1694839300 residues

Total number of hits satisfying chosen parameters: 4770830

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters: -DB=exlh

-MODE=frame+p2n.model -DB=exlh  
-Q=/cgn2\_1/USPTO.spool/US09847539/runat\_10102002\_092550\_3509/app\_query.fasta\_1.526  
-DB=Pending\_Patents\_NA\_New -OFMT=fastap -SUFFIX=trnp -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=40  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09847539 -CGN\_1\_1.186 -runat\_10102002\_092550\_3509 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FCAPOP=6 -FCAPEXT=7  
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

Pending Patents\_NA\_New:\*  
1: /cgn2\_6/ptodata/1/pna/PCT\_NEW\_COMB.seq:\*  
2: /cgn2\_6/ptodata/1/pna/US06\_NEW\_COMB.seq:\*  
3: /cgn2\_6/ptodata/1/pna/US07\_NEW\_COMB.seq:\*  
4: /cgn2\_6/ptodata/1/pna/US08\_NEW\_COMB.seq:\*  
5: /cgn2\_6/ptodata/1/pna/US09\_NEW\_COMB.seq:\*  
6: /cgn2\_6/ptodata/1/pna/US10\_NEW\_COMB.seq:\*  
7: /cgn2\_6/ptodata/1/pna/US10\_NEW\_COMB.seq2:\*  
8: /cgn2\_6/ptodata/1/pna/US60\_NEW\_COMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	50	38.8	1679	US-10-219-999-12394	Sequence 12394, A
C 2	50	38.8	1896	US-10-219-999-8111	Sequence 8111, Ap
C 3	50	38.8	1783	US-10-155-881-3337	Sequence 3337, Ap
C 4	50	38.8	3257	US-60-360-039-30630	Sequence 30630, A
C 5	48	37.2	1449	US-60-360-039-40312	Sequence 40312, A
C 6	48	37.2	1449	US-60-360-039-40312	Sequence 40312, A
C 7	47	36.4	783	US-10-179-131-3446	Sequence 3446, Ap
C 8	47	36.4	4885	US-60-360-039-40814	Sequence 40814, A
C 9	47	36.4	6965	US-60-360-039-27525	Sequence 27525, A
C 10	47	36.4	59065	US-09-440-302B-636	Sequence 636, App
				US-10-135-696-3	Sequence 3, Appli

11	47	36.4	128978	1	PCT-US01-17423-6	Sequence 6, Appli
12	47	36.4	128978	1	PCT-US02-25766-4301	Sequence 4301, Ap
13	47	36.4	128978	7	US-10-161-453-6	Sequence 6, Appli
C 14	46	35.7	705	7	US-10-155-881-3217	Sequence 3217, Ap
C 15	46	35.7	1512	6	US-10-219-999-6030	Sequence 6030, Ap
C 16	46	35.7	1608	8	US-60-360-039-44539	Sequence 44539, A
C 17	46	35.7	2199	8	US-60-360-039-42394	Sequence 42394, A
C 18	45	34.9	825	7	US-10-027-632-164206	Sequence 164206, A
C 19	45	34.9	825	7	US-10-027-632-164207	Sequence 164207, A
C 20	45	34.9	825	7	US-10-027-632-164208	Sequence 164208, A
C 21	45	34.9	950	6	US-10-219-999-22677	Sequence 22677, A
C 22	45	34.9	1557	8	US-60-360-039-27473	Sequence 27473, A
C 23	45	34.9	2243	7	US-10-027-632-262301	Sequence 262301, A
C 24	45	34.9	2688	8	US-60-360-039-28506	Sequence 28506, A
C 25	45	34.9	2845	8	US-60-360-039-31265	Sequence 31265, A
C 26	45	34.9	3435	8	US-60-360-039-44171	Sequence 44171, A
C 27	44	34.1	237	5	US-09-975-254-11911	Sequence 11911, A
C 28	44	34.1	370	5	US-09-918-995-37583	Sequence 37583, A
C 29	44	34.1	373	5	US-09-918-995-37603	Sequence 37603, A
C 30	44	34.1	425	6	US-10-217-489-137	Sequence 137, App
C 31	44	34.1	489	6	US-10-221-279-4262	Sequence 4262, Ap
C 32	44	34.1	511	7	US-10-027-632-13710	Sequence 13710, A
C 33	44	34.1	546	5	US-09-918-995-23409	Sequence 23409, A
C 34	44	34.1	656	1	PCT-US02-27884-4854	Sequence 4854, Ap
C 35	44	34.1	656	1	PCT-US02-27884-5555	Sequence 5555, Ap
C 36	44	34.1	862	7	US-10-027-632-172753	Sequence 172753, A
C 37	44	34.1	862	7	US-10-027-632-172754	Sequence 172754, A
C 38	44	34.1	862	7	US-10-027-632-172755	Sequence 172755, A
C 39	44	34.1	940	7	US-10-155-881-5807	Sequence 5807, Ap
C 40	44	34.1	1033	4	US-08-961-083-191	Sequence 191, App
C 41	44	34.1	1216	6	US-10-217-550-124	Sequence 124, App
C 42	44	34.1	1428	7	US-10-155-881-17609	Sequence 17609, A
C 43	44	34.1	1484	6	US-10-219-999-26942	Sequence 26942, A
C 44	44	34.1	1811	6	US-10-219-999-30708	Sequence 30708, A
C 45	44	34.1	1890	8	US-60-360-039-44420	Sequence 44420, A

#### ALIGNMENTS

RESULT 1  
US-10-219-999-12394/c  
; Sequence 12394, Application US/10219999  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Edgerton, Michael D  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Liu, Jingdong  
; APPLICANT: Stein, Joshua  
; TITLE OF INVENTION: CDNA SEQUENCES AND USES FOR PLANT IMPROVEMENT  
; FILE REFERENCE: 38-10(52726)C  
; CURRENT APPLICATION NUMBER: US/10/219,999  
; CURRENT FILING DATE: 2002-08-15  
; PRIOR APPLICATION NUMBER: US 60/324,109  
; PRIOR FILING DATE: 2001-09-21  
; PRIOR APPLICATION NUMBER: US 60/312,544  
; PRIOR FILING DATE: 2001-08-15  
; NUMBER OF SEQ ID NOS: 63520  
; SEQ ID NO 12394  
; LENGTH: 1679  
; TYPE: DNA  
; ORGANISM: Zea mays  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (2)..(1351)  
; OTHER INFORMATION:  
US-10-219-999-12394  
Alignment Scores: 12.8 Length: 1679  
Pred. No.: 50.00 Matches: 11  
Score: 73.08% Conservative: 8  
Percent Similarity: 42.31% Mismatches: 7  
Best Local Similarity:

**THIS PAGE BLANK (USPTO)**

```
; NAME/KEY: misc_feature
; LOCATION: (1)..(1230)
; OTHER INFORMATION: any n = a, g, c, t, unknown, or other
; NAME/KEY: misc_feature
; LOCATION: (1)..(1230)
; OTHER INFORMATION: 40182 : FINISHED (Clone Number : FINISHED)
US-09-565-309A-62025
```

```
Alignment Scores:
Pred. No.: 177 Length: 1230
Score: 52.00 Matches: 11
Percent Similarity: 72.73% Conservative: 5
Best Local Similarity: 50.00% Mismatches: 6
Query Match: 40.31% Indels: 0
DB: 22 Gaps: 0
```

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-09-565-309A-62025 (1-1230)

```
Qy 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVallys 23
||| |||||:||||: ||| ||| :|||:|||||: |||
Db 518 CTTCCGGCACTCCGACAAAGGACATGCTCAAAATTCAAAAGCTCAATCGCAAG 459

Qy 24 AlaAsp 25
|||||
Db 458 GCGGAT 453
```

```
RESULT 40
US-09-595-329A-545/c
; Sequence 545, Application US/09595329A
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nikolai
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptide
; FILE REFERENCE: 2750-0948P
; CURRENT APPLICATION NUMBER: US/09/595,329A
; CURRENT FILING DATE: 2000-06-16
; NUMBER OF SEQ ID NOS: 3309
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 545
; LENGTH: 1230
; TYPE: DNA
; ORGANISM: Arabidopsis Thaliana
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(1230)
; OTHER INFORMATION: Ceres Seq. ID no. 1013068
; NAME/KEY: misc_feature
; LOCATION: (1)..(1)
; OTHER INFORMATION: n is a, c, t, g, unknown, or other
US-09-595-329A-545
```

```
Alignment Scores:
Pred. No.: 177 Length: 1230
Score: 52.00 Matches: 11
Percent Similarity: 72.73% Conservative: 5
Best Local Similarity: 50.00% Mismatches: 6
Query Match: 40.31% Indels: 0
DB: 22 Gaps: 0
```

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-09-595-329A-545 (1-1230)

```
Qy 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVallys 23
||| |||||:||||: ||| ||| :|||:|||||: |||
Db 518 CTTCCGGCACTCCGACAAAGGACATGCTCAAAATTCAAAAGCTCAATCGCAAG 459

Qy 24 AlaAsp 25
|||||
Db 458 GCGGAT 453
```

Search completed: October 13, 2002, 07:14:46  
Job time : 567.553 secs



```
RESULT 36
US-09-565-309A-67148/c
; Sequence 67148, Application US/09565309A
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nickolai
; APPLICANT: BROVER, Vyacheslav
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE REFERENCE: 2750-0853P
; CURRENT APPLICATION NUMBER: US/09/565,309A
; CURRENT FILING DATE: 2000-05-05
; NUMBER OF SEQ ID NOS: 68449
; SEQ ID NO 67148
; LENGTH: 939
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(939)
; OTHER INFORMATION: any n = a, g, c, t, unknown, or other
; NAME/KEY: misc_feature
; LOCATION: (1)..(939)
; OTHER INFORMATION: 40182 : CDS (Clone Number : CDS)
US-09-565-309A-67148

Alignment Scores:
Pred. No.: 130 Length: 939
Score: 52.00 Matches: 11
Percent Similarity: 72.73% Conservative: 5
Best Local Similarity: 50.00% Mismatches: 6
Query Match: 40.31% Indels: 0
DB: 22 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x US-09-565-309A-67148 (1-939)
QY 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVallys 23
DB 382 CTTCGCGCACTCTCCGACAAAGGACATGCCTCAAAATTCAAAAAGCTGCAATCGCAAG 323

QY 24 AlaAsp 25
DB 322 GCGGAT 317

RESULT 37
US-09-708-427-29368/c
; Sequence 29368, Application US/09708427
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE REFERENCE: 2750-1243P
; CURRENT APPLICATION NUMBER: US/09/708,427
; CURRENT FILING DATE: 2000-11-09
; NUMBER OF SEQ ID NOS: 85364
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 29368
; LENGTH: 984
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..984
; OTHER INFORMATION: any n = a, g, c, t, unknown, or other
; NAME/KEY: misc_feature
; LOCATION: 1..984
; OTHER INFORMATION: Ceres Seq. ID 1825505
US-09-708-427-29368

Alignment Scores:
Pred. No.: 137 Length: 984
Score: 52.00 Matches: 11
Percent Similarity: 72.73% Conservative: 5
Best Local Similarity: 50.00% Mismatches: 6
```

```
Query Match: 40.31% Indels: 0
DB: 28 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x US-09-708-427-29368 (1-984)
QY 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVallys 23
DB 442 CTTCGCGCACTCTCCGACAAAGGACATGCCTCAAAATTCAAAAAGCTGCAATCGCAAG 383

QY 24 AlaAsp 25
DB 382 GCGGAT 377

RESULT 38
US-09-565-309A-67147/c
; Sequence 67147, Application US/09565309A
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nickolai
; APPLICANT: BROVER, Vyacheslav
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTID
; FILE REFERENCE: 2750-0853P
; CURRENT APPLICATION NUMBER: US/09/565,309A
; CURRENT FILING DATE: 2000-05-05
; NUMBER OF SEQ ID NOS: 68449
; SEQ ID NO 67147
; LENGTH: 999
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(999)
; OTHER INFORMATION: any n = a, g, c, t, unknown, or other
; NAME/KEY: misc_feature
; LOCATION: (1)..(999)
; OTHER INFORMATION: 40182 : CDS (Clone Number : CDS)
US-09-565-309A-67147

Alignment Scores:
Pred. No.: 139 Length: 999
Score: 52.00 Matches: 11
Percent Similarity: 72.73% Conservative: 5
Best Local Similarity: 50.00% Mismatches: 6
Query Match: 40.31% Indels: 0
DB: 22 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x US-09-565-309A-67147 (1-999)
QY 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVallys 23
DB 442 CTTCGCGCACTCTCCGACAAAGGACATGCCTCAAAATTCAAAAAGCTGCAATCGCAAG 383

QY 24 AlaAsp 25
DB 382 GCGGAT 377

RESULT 39
US-09-565-309A-62025/c
; Sequence 62025, Application US/09565309A
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nickolai
; APPLICANT: BROVER, Vyacheslav
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTID
; FILE REFERENCE: 2750-0853P
; CURRENT APPLICATION NUMBER: US/09/565,309A
; CURRENT FILING DATE: 2000-05-05
; NUMBER OF SEQ ID NOS: 68449
; SEQ ID NO 62025
; LENGTH: 1230
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
```

```
; APPLICANT: RUSSO, FRANK D.
; APPLICANT: HANN, AMY L.
; APPLICANT: HEATH, JOE D.
; APPLICANT: FINNEY, GREGORY L.
; APPLICANT: BROOKS, JACQUELINE
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES OF PSEUDOMONAS AERUGINOSA
; TITLE OF INVENTION: GENOME, FRAGMENTS THEREOF, AND USES THEREOF
; NUMBER OF SEQUENCES: 860
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/60/082,302
; FILING DATE: HEREWITH
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: CERRONE, MICHAEL C.
; REGISTRATION NUMBER: 39,132
; REFERENCE/DOCKET NUMBER: PM-0011-1 P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 513:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8830 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: genomic DNA
; IMMEDIATE SOURCE:
; CLONE: PAELC517
; US-60-082-302-513

Alignment Scores:
Pred. No.: 1,41e+03 Length: 8830
Score: 52.50 Matches: 14
Percent Similarity: 71.43% Conservative: 6
Best Local Similarity: 50.00% Mismatches: 3
Query Match: 40.70% Indels: 5
DB: 47 Gaps: 1

US-09-847-539A-6_COPY_59_86 (1-28) x US-60-082-302-513 (1-8830)
Qy 5 GluAlaLeuAlaAsp-----GlnThrAspAlaLeuGlnSerGluGluAla 19
|||||:|||||:
Db 2929 GAAGCGTAGTGTAGCGTACGACAGCGGAGCGGACTGCGTGTGCGAGCGGATGTCGCC 2870
|||||:|||||:

Qy 20 AlaValVallysAlaAspAsnAla 27
|||||:|||||:
Db 2869 GCTGTGCTAGGCGCCGACCTGCGC 2846
|||||:|||||:

RESULT 34
US-09-565-309A-32821/C
; Sequence 32821, Application US/09565309A
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nikolai
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; TITLE OF INVENTION: THEREBY
; FILE REFERENCE: 2750-0853P
; CURRENT APPLICATION NUMBER: US/09/565,309A
; CURRENT FILING DATE: 2000-05-05
; NUMBER OF SEQ ID NOS: 68449
; SEQ ID NO 32821
```

```
; LENGTH: 594
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(594)
; OTHER INFORMATION: any n = a, g, c, t, unknown, or other
; NAME/KEY: misc_feature
; LOCATION: (1)..(594)
; OTHER INFORMATION: 40182:107311 (Clone Number:Unique Sequence Identifier)
US-09-565-309A-32821

Alignment Scores:
Pred. No.: 76.4 Length: 594
Score: 52.00 Matches: 11
Percent Similarity: 72.73% Conservative: 5
Best Local Similarity: 50.00% Mismatches: 6
Query Match: 40.31% Indels: 0
DB: 22 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x US-09-565-309A-32821 (1-594)
Qy 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVallys 23
|||||:|||||:|||||:
Db 518 CTTCGCGACTCTCCGACAAAGGACATGCTCAAAATTCAAAAGCTGCAATGCCAAAG 459
|||||:|||||:

Qy 24 AlaAsp 25
|||||:
Db 458 GCGGAT 453

RESULT 35
US-09-565-309A-32824/C
; Sequence 32824, Application US/09565309A
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nikolai
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTID
; TITLE OF INVENTION: THEREBY
; FILE REFERENCE: 2750-0853P
; CURRENT APPLICATION NUMBER: US/09/565,309A
; CURRENT FILING DATE: 2000-05-05
; NUMBER OF SEQ ID NOS: 68449
; SEQ ID NO 32824
; LENGTH: 641
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(641)
; OTHER INFORMATION: any n = a, g, c, t, unknown, or other
; NAME/KEY: misc_feature
; LOCATION: (1)..(641)
; OTHER INFORMATION: 40182:957281 (Clone Number:Unique Sequence Identifier)
US-09-565-309A-32824

Alignment Scores:
Pred. No.: 83.4 Length: 641
Score: 52.00 Matches: 11
Percent Similarity: 72.73% Conservative: 5
Best Local Similarity: 50.00% Mismatches: 6
Query Match: 40.31% Indels: 0
DB: 22 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x US-09-565-309A-32824 (1-641)
Qy 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVallys 23
|||||:|||||:|||||:
Db 461 CTTCGCGACTCTCCGACAAAGGACATGCTCAAAATTCAAAAGCTGCAATGCCAAAG 402
|||||:|||||:

Qy 24 AlaAsp 25
|||||:
Db 401 GCGGAT 396
```

; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 17589  
; LENGTH: 7206  
; TYPE: DNA  
; ORGANISM: Drosophila  
US-60-167-217-17589

Alignment Scores:  
Pred. No.: 619 Length: 7206  
Score: 54.00 Matches: 12  
Percent Similarity: 66.67% Conservative: 4  
Best Local Similarity: 50.00% Mismatches: 8  
Query Match: 41.86% Indels: 0  
DB: 55 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-60-167-217-17589 (1-7206)

QY 5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaValValLysAla 24  
DB 4905 GAGGAGGAAGCTGACGAGGAGGAGCGCGCGGAGGATGCCGCGCGGCTGCCGCC 4964  
QY 25 AspAsnAlaAla 28  
DB 4965 GATGCTGCGCGC 4976

RESULT 30

US-09-528-237A-1811/c  
; Sequence 1811, Application US/09528237A

; GENERAL INFORMATION:  
; APPLICANT: Venter, J. Craig  
; TITLE OF INVENTION: The Drosophila Genome, Primary Nucleic  
; TITLE OF INVENTION: Acid Sequences, Systems Containing The Nucleic Acid  
; FILE OF INVENTION: Sequences and Uses Thereof  
; FILE REFERENCE: CLO00284  
; CURRENT APPLICATION NUMBER: US/09/528,237A  
; CURRENT FILING DATE: 2000-03-17  
; NUMBER OF SEQ ID NOS: 2926

; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1811  
; LENGTH: 32275  
; TYPE: DNA  
; ORGANISM: Drosophila  
US-09-528-237A-1811

Alignment Scores:  
Pred. No.: 3,48e+03 Length: 32275  
Score: 54.00 Matches: 12  
Percent Similarity: 66.67% Conservative: 4  
Best Local Similarity: 50.00% Mismatches: 8  
Query Match: 41.86% Indels: 0  
DB: 19 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-09-528-237A-1811 (1-32275)

QY 5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaValValLysAla 24  
DB 26995 GAGGAGGAAGCTGACGAGGAGGAGCGCGCGGAGGATGCCGCGCGGCTGCCGCC 26936  
QY 25 AspAsnAlaAla 28  
DB 26935 GATGCTGCGCGC 26924

RESULT 31

US-09-252-991A-8951/c  
; Sequence 8951, Application US/09252991A

; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 8951  
; LENGTH: 1239  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-8951

Alignment Scores:  
Pred. No.: 147 Length: 1239  
Score: 52.50 Matches: 14  
Percent Similarity: 71.43% Conservative: 6  
Best Local Similarity: 50.00% Mismatches: 3  
Query Match: 40.70% Indels: 5  
DB: 16 Gaps: 1

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-09-252-991A-8951 (1-1239)

QY 5 GluAlaLeuAlaAsp-----GlnThrAspAlaLeuGlnSerGluGluAla 19  
DB 466 GAAGCGGTAGCTGAGGCTGAGCAGCGGACGACGCTGTGCGAGCGCGATGCGCC 407  
QY 20 AlaValValLysAlaAspAsnAla 27  
DB 406 GCTGCTGCTAGGCGCGCACCTGCGC 383

RESULT 32

US-09-252-991A-8732/c

; Sequence 8732, Application US/09252991A  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 8732  
; LENGTH: 2127  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-8732

Alignment Scores:  
Pred. No.: 273 Length: 2127  
Score: 52.50 Matches: 14  
Percent Similarity: 71.43% Conservative: 6  
Best Local Similarity: 50.00% Mismatches: 3  
Query Match: 40.70% Indels: 5  
DB: 16 Gaps: 1

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-09-252-991A-8732 (1-2127)

QY 5 GluAlaLeuAlaAsp-----GlnThrAspAlaLeuGlnSerGluGluAla 19  
DB 1608 GAAGCGGTAGCTGAGGCTGAGCAGCGGACGACGCTGTGCGAGCGCGATGCGCC 1549  
QY 20 AlaValValLysAlaAspAsnAla 27  
DB 1548 GCTGCTGCTAGGCGCGCACCTGCGC 1525

RESULT 33

US-60-082-302-513/c

; Sequence 513, Application US/60082302  
; GENERAL INFORMATION:  
; APPLICANT: LAGACE, ROBERT E.  
; APPLICANT: CORLEY, NEIL C.

```
; LENGTH: 6398
; TYPE: DNA
; ORGANISM: DROSOPHILA
US-09-614-150-17530

Alignment Scores:
Pred. No.: 539          Length: 6398
Score: 54.00           Matches: 12
Percent Similarity: 66.67% Conservative: 4
Best Local Similarity: 50.00% Mismatches: 8
Query Match: 41.86%      Indels: 0
DB: 23                 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x US-09-614-150-17530 (1-6398)
Qy 5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVallysAla 24
   ||| |||||: ||||| :|||: ||||| |||
Db 2298 GAGGAGGAAGCTGACGAGGAGGACGCCGCCGAGGAGGATGCCGCGCTGCCGCC 2239
Qy 25 AspAsnAlaAla 28
   ||| |||||
Db 2238 GATGCTGCGGCG 2227

RESULT 26
US-60-191-637-17580/c
; Sequence 17580, Application US/60191637
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; TITLE OF INVENTION: NUCLEIC ACID DETECTION KITS COMPRISING
; TITLE OF INVENTION: GENE SEQUENCES EXPRESSED FROM THE DROSOPHILA GENOME, AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL000392
; CURRENT APPLICATION NUMBER: US/60/191,637
; CURRENT FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 42660
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17580
; LENGTH: 6398
; TYPE: DNA
; ORGANISM: DROSOPHILA
US-60-191-637-17580

Alignment Scores:
Pred. No.: 539          Length: 6398
Score: 54.00           Matches: 12
Percent Similarity: 66.67% Conservative: 4
Best Local Similarity: 50.00% Mismatches: 8
Query Match: 41.86%      Indels: 0
DB: 58                 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x US-60-191-637-17580 (1-6398)
Qy 5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVallysAla 24
   ||| |||||: ||||| :|||: ||||| |||
Db 2298 GAGGAGGAAGCTGACGAGGAGGACGCCGCCGAGGAGGATGCCGCGCTGCCGCC 2239
Qy 25 AspAsnAlaAla 28
   ||| |||||
Db 2238 GATGCTGCGGCG 2227

RESULT 27
US-60-191-681-13896/c
; Sequence 13896, Application US/60191681
; GENERAL INFORMATION:
; APPLICANT: Li, Peter, W.D.
; TITLE OF INVENTION: ISOLATED DROSOPHILA PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING ISOLATED DROSOPHILA PROTEINS AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL000390
; CURRENT APPLICATION NUMBER: US/60/191,681
; CURRENT FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 30973
; SOFTWARE: FastSEQ for Windows Version 4.0
```

```
; SEQ ID NO 13896
; LENGTH: 6398
; TYPE: DNA
; ORGANISM: DROSOPHILA
US-60-191-681-13896

Alignment Scores:
Pred. No.: 539          Length: 6398
Score: 54.00           Matches: 12
Percent Similarity: 66.67% Conservative: 4
Best Local Similarity: 50.00% Mismatches: 8
Query Match: 41.86%      Indels: 0
DB: 58                 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x US-60-191-681-13896 (1-6398)
Qy 5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVallysAla 24
   ||| |||||: ||||| :|||: ||||| |||
Db 2298 GAGGAGGAAGCTGACGAGGAGGACGCCGCCGAGGAGGATGCCGCGCTGCCGCC 2239
Qy 25 AspAsnAlaAla 28
   ||| |||||
Db 2238 GATGCTGCGGCG 2227

RESULT 28
US-60-173-464-14400
; Sequence 14400, Application US/60173464
; GENERAL INFORMATION:
; APPLICANT: Li, Peter W.D.
; TITLE OF INVENTION: ISOLATED G-PROTEIN COUPLED RECEPTORS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING GPCR PROTEINS AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL000173
; CURRENT APPLICATION NUMBER: US/60/173,464
; CURRENT FILING DATE: 1999-12-29
; NUMBER OF SEQ ID NOS: 30269
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14400
; LENGTH: 7203
; TYPE: DNA
; ORGANISM: Drosophila
US-60-173-464-14400

Alignment Scores:
Pred. No.: 618          Length: 7203
Score: 54.00           Matches: 12
Percent Similarity: 66.67% Conservative: 4
Best Local Similarity: 50.00% Mismatches: 8
Query Match: 41.86%      Indels: 0
DB: 56                 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x US-60-173-464-14400 (1-7203)
Qy 5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVallysAla 24
   ||| |||||: ||||| :|||: ||||| |||
Db 4905 GAGGAGGAAGCTGACGAGGAGGACGCCGCCGAGGAGGATGCCGCGCTGCCGCC 4964
Qy 25 AspAsnAlaAla 28
   ||| |||||
Db 4965 GATGCTGCGGCG 4976

RESULT 29
US-60-167-217-17589
; Sequence 17589, Application US/60167217
; GENERAL INFORMATION:
; APPLICANT: Li, Peter W. D.
; TITLE OF INVENTION: ISOLATED DROSOPHILA PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING DROSOPHILA PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL000152
; CURRENT APPLICATION NUMBER: US/60/167,217
; CURRENT FILING DATE: 1999-11-24
; NUMBER OF SEQ ID NOS: 23195
```

; PRIOR FILING DATE: 1999-10-05  
; PRIOR APPLICATION NUMBER: 60/160,191  
; PRIOR FILING DATE: 1999-10-19  
; PRIOR APPLICATION NUMBER: 60/161,932  
; PRIOR FILING DATE: 1999-10-28  
; PRIOR APPLICATION NUMBER: 60/164,769  
; PRIOR FILING DATE: 1999-11-12  
; PRIOR APPLICATION NUMBER: 60/173,383  
; PRIOR FILING DATE: 1999-12-28  
; PRIOR APPLICATION NUMBER: 60/175,693  
; PRIOR FILING DATE: 2000-01-12  
; PRIOR APPLICATION NUMBER: 60/184,831  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: 60/191,637  
; PRIOR FILING DATE: 2000-03-23  
; NUMBER OF SEQ ID NOS: 43008  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 17531  
; LENGTH: 3349  
; TYPE: DNA  
; ORGANISM: DROSOPHILA  
US-09-614-150-17531

Alignment Scores:  
Pred. No.: 256 Length: 3349  
Score: 54.00 Matches: 12  
Percent Similarity: 66.67% Conservative: 4  
Best Local Similarity: 50.00% Mismatches: 8  
Query Match: 41.86% Indels: 0  
DB: 23 Gaps: 0

US-09-847-539a-6\_COPY\_59\_86 (1-28) x US-09-614-150-17531 (1-3349)

QY 5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaValVallysAla 24  
||| |||||::: ||||| :::::||||:||||| |||||  
Db 1122 GAGGAGGAAGCTGACGAGGAGGACGCCGCCGAGGAGGATGCCGCGCGCTGCCGCC 1063

QY 25 AspAsnAlaAla 28  
||| |||||  
Db 1062 GATGCTGCGCG 1051

RESULT 23

US-60-191-637-17581/c

; Sequence 17581, Application US/60191637  
; GENERAL INFORMATION:  
; APPLICANT: Venter, J. Craig  
; TITLE OF INVENTION: NUCLEIC ACID DETECTION KITS COMPRISING  
; TITLE OF INVENTION: GENE SEQUENCES EXPRESSED FROM THE DROSOPHILA GENOME, AND  
; TITLE OF INVENTION: USES THEREOF  
; FILE REFERENCE: CL000392  
; CURRENT APPLICATION NUMBER: US/60/191,637  
; CURRENT FILING DATE: 2000-03-23  
; NUMBER OF SEQ ID NOS: 42660  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 17581  
; LENGTH: 3349  
; TYPE: DNA  
; ORGANISM: DROSOPHILA  
US-60-191-637-17581

Alignment Scores:  
Pred. No.: 256 Length: 3349  
Score: 54.00 Matches: 12  
Percent Similarity: 66.67% Conservative: 4  
Best Local Similarity: 50.00% Mismatches: 8  
Query Match: 41.86% Indels: 0  
DB: 58 Gaps: 0

US-09-847-539a-6\_COPY\_59\_86 (1-28) x US-60-191-637-17581 (1-3349)

QY 5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaValVallysAla 24  
||| |||||::: ||||| :::::||||:||||| |||||  
Db 1122 GAGGAGGAAGCTGACGAGGAGGACGCCGCCGAGGAGGATGCCGCGCGCTGCCGCC 1063

QY 25 AspAsnAlaAla 28  
||| |||||  
Db 1062 GATGCTGCGCG 1051

RESULT 24

US-60-191-681-13897/c

; Sequence 13897, Application US/60191681  
; GENERAL INFORMATION:  
; APPLICANT: Li, Peter, W.D.  
; TITLE OF INVENTION: ISOLATED DROSOPHILA PROTEINS, NUCLEIC  
; TITLE OF INVENTION: ACID MOLECULES ENCODING ISOLATED DROSOPHILA PROTEINS AND  
; TITLE OF INVENTION: USES THEREOF.  
; FILE REFERENCE: CL000390  
; CURRENT APPLICATION NUMBER: US/60/191,681  
; CURRENT FILING DATE: 2000-03-23  
; NUMBER OF SEQ ID NOS: 30973  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 13897  
; LENGTH: 3349  
; TYPE: DNA  
; ORGANISM: DROSOPHILA  
US-60-191-681-13897

Alignment Scores:  
Pred. No.: 256 Length: 3349  
Score: 54.00 Matches: 12  
Percent Similarity: 66.67% Conservative: 4  
Best Local Similarity: 50.00% Mismatches: 8  
Query Match: 41.86% Indels: 0  
DB: 58 Gaps: 0

US-09-847-539a-6\_COPY\_59\_86 (1-28) x US-60-191-681-13897 (1-3349)

QY 5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaValVallysAla 24  
||| |||||::: ||||| :::::||||:||||| |||||  
Db 1122 GAGGAGGAAGCTGACGAGGAGGACGCCGCCGAGGAGGATGCCGCGCGCTGCCGCC 1063

QY 25 AspAsnAlaAla 28  
||| |||||  
Db 1062 GATGCTGCGCG 1051

RESULT 25

US-09-614-150-17530/c

; Sequence 17530, Application US/09614150  
; GENERAL INFORMATION:  
; APPLICANT: Venter, J. Craig  
; APPLICANT: et al.  
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID  
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE  
; TITLE OF INVENTION: DROSOPHILA GENES.  
; FILE REFERENCE: CL000728  
; CURRENT APPLICATION NUMBER: US/09/614,150  
; CURRENT FILING DATE: 2000-07-11  
; PRIOR APPLICATION NUMBER: 60/157,832  
; PRIOR FILING DATE: 1999-10-05  
; PRIOR APPLICATION NUMBER: 60/160,191  
; PRIOR FILING DATE: 1999-10-19  
; PRIOR APPLICATION NUMBER: 60/161,932  
; PRIOR FILING DATE: 1999-10-28  
; PRIOR APPLICATION NUMBER: 60/164,769  
; PRIOR FILING DATE: 1999-11-12  
; PRIOR APPLICATION NUMBER: 60/173,383  
; PRIOR FILING DATE: 1999-12-28  
; PRIOR APPLICATION NUMBER: 60/175,693  
; PRIOR FILING DATE: 2000-01-12  
; PRIOR APPLICATION NUMBER: 60/184,831  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: 60/191,637  
; PRIOR FILING DATE: 2000-03-23  
; NUMBER OF SEQ ID NOS: 43008  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 17530

```
; LENGTH: 1783
; TYPE: DNA
; ORGANISM: Drosophila
US-60-142-845-289

Alignment Scores:
Pred. No.: 124      Length: 1783
Score: 54.00      Matches: 12
Percent Similarity: 66.67%      Conservative: 4
Best Local Similarity: 50.00%      Mismatches: 8
Query Match: 41.86%      Indels: 0
DB: 53      Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x US-60-142-845-289 (1-1785)

Qy 5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLysAla 24
   ||| |||||: ||||| :|||: ||||| |||
Db 1475 GAGGAGGAAGCTGACGAGGAGGCGCGGCGGAGGATGCGGCGGCTGCCGCC 1534

Qy 25 AspAsnAlaAla 28
   ||| |||||
Db 1535 GATGCTCGGCGG 1546

RESULT 19
US-60-145-134-143
; Sequence 143, Application US/60145134
; GENERAL INFORMATION:
; APPLICANT: Bonazzi, Vivien
; TITLE OF INVENTION: ISOLATED DROSOPHILA PROTEINS, NUCLEIC
; FILE REFERENCE: CL000058
; CURRENT APPLICATION NUMBER: US/60/145,134
; CURRENT FILING DATE: 1999-07-22
; NUMBER OF SEQ ID NOS: 342
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 143
; LENGTH: 1785
; TYPE: DNA
; ORGANISM: Drosophila
US-60-145-134-143

Alignment Scores:
Pred. No.: 124      Length: 1785
Score: 54.00      Matches: 12
Percent Similarity: 66.67%      Conservative: 4
Best Local Similarity: 50.00%      Mismatches: 8
Query Match: 41.86%      Indels: 0
DB: 53      Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x US-60-145-134-143 (1-1785)

Qy 5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLysAla 24
   ||| |||||: ||||| :|||: ||||| |||
Db 1477 GAGGAGGAAGCTGACGAGGAGGCGCGGCGGAGGATGCGGCGGCTGCCGCC 1536

Qy 25 AspAsnAlaAla 28
   ||| |||||
Db 1537 GATGCTCGGCGG 1548

RESULT 20
US-60-173-464-14401/c
; Sequence 14401, Application US/60173464
; GENERAL INFORMATION:
; APPLICANT: Li, Peter W.D.
; TITLE OF INVENTION: ISOLATED G-PROTEIN COUPLED RECEPTORS,
; FILE REFERENCE: CL000173
; CURRENT APPLICATION NUMBER: US/60/173,464
; CURRENT FILING DATE: 1999-12-29
; NUMBER OF SEQ ID NOS: 30269
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14401
```

```
; LENGTH: 3046
; TYPE: DNA
; ORGANISM: Drosophila
US-60-173-464-14401

Alignment Scores:
Pred. No.: 229      Length: 3046
Score: 54.00      Matches: 12
Percent Similarity: 66.67%      Conservative: 4
Best Local Similarity: 50.00%      Mismatches: 8
Query Match: 41.86%      Indels: 0
DB: 56      Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x US-60-173-464-14401 (1-3046)

Qy 5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLysAla 24
   ||| |||||: ||||| :|||: ||||| |||
Db 1014 GAGGAGGAAGCTGACGAGGAGGCGCGGCGGAGGATGCGGCGGCTGCCGCC 955

Qy 25 AspAsnAlaAla 28
   ||| |||||
Db 954 GATGCTCGGCGG 943

RESULT 21
US-60-167-217-17590/c
; Sequence 17590, Application US/60167217
; GENERAL INFORMATION:
; APPLICANT: Li, Peter W. D.
; TITLE OF INVENTION: ISOLATED DROSOPHILA PROTEINS, NUCLEIC
; FILE REFERENCE: CL000152
; CURRENT APPLICATION NUMBER: US/60/167,217
; CURRENT FILING DATE: 1999-11-24
; NUMBER OF SEQ ID NOS: 23195
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17590
; LENGTH: 3049
; TYPE: DNA
; ORGANISM: Drosophila
US-60-167-217-17590

Alignment Scores:
Pred. No.: 230      Length: 3049
Score: 54.00      Matches: 12
Percent Similarity: 66.67%      Conservative: 4
Best Local Similarity: 50.00%      Mismatches: 8
Query Match: 41.86%      Indels: 0
DB: 55      Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x US-60-167-217-17590 (1-3049)

Qy 5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLysAla 24
   ||| |||||: ||||| :|||: ||||| |||
Db 1017 GAGGAGGAAGCTGACGAGGAGGCGCGGCGGAGGATGCGGCGGCTGCCGCC 958

Qy 25 AspAsnAlaAla 28
   ||| |||||
Db 957 GATGCTCGGCGG 946

RESULT 22
US-09-614-150-17531/c
; Sequence 17531, Application US/09614150
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/09/614,150
; CURRENT FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/157,832
```

```
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 10031
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: Windows NT 4.0
; SOFTWARE: UNIX
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/993,002A
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: GTN-018
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)742-4214
; INFORMATION FOR SEQ ID NO: 4235:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1239 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 1...1239
;
US-08-993-002A-4235
Alignment Scores:
Pred. No.: 25.1 Length: 1239
Score: 57.00 Matches: 12
Percent Similarity: 69.57% Conservative: 4
Best Local Similarity: 52.17% Mismatches: 7
Query Match: 44.19% Indels: 0
DB: 13 Gaps: 0

US-09-847-539a-6_COPY_59_86 (1-28) x US-08-993-002A-4235 (1-1239)
Qy 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLys 23
Db 145 ATACAAGCCCTACAGGACAAATTGACGCTTTAGATTCTCAAGAAAAAGTCGTTAGCAA 204
Qy 24 AlaAspAsn 26
Db 205 TGGGATAAC 213
RESULT 17
US-08-621-425-49
; Sequence 49, Application US/08621425
; GENERAL INFORMATION:
; APPLICANT: Douglas Smith
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; TITLE OF INVENTION: RELATING TO HELICOBACTER PYLORI FOR
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 495
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street, Suite 510
```

```
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 8mm cartridge tape
; COMPUTER: SPARC station LX
; OPERATING SYSTEM: sunOS4
; SOFTWARE: tar
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/621,425
; FILING DATE: 25-MARCH-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/561,469
; FILING DATE: 17-NOV-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/487,032
; FILING DATE: 07-JUNE-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: GTN-001CP2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 49:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3497 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 1...3497
;
US-08-621-425-49
Alignment Scores:
Pred. No.: 83 Length: 3497
Score: 57.00 Matches: 12
Percent Similarity: 69.57% Conservative: 4
Best Local Similarity: 52.17% Mismatches: 7
Query Match: 44.19% Indels: 0
DB: 10 Gaps: 0

US-09-847-539a-6_COPY_59_86 (1-28) x US-08-621-425-49 (1-3497)
Qy 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLys 23
Db 1528 ATACAAGCCCTACAGGACAAATTGACGCTTTAGATTCTCAAGAAAAAGTCGTTAGCAA 1587
Qy 24 AlaAspAsn 26
Db 1588 TGGGATAAC 1596
RESULT 18
US-60-142-845-289
; Sequence 289, Application US/60142845
; GENERAL INFORMATION:
; APPLICANT: Kerlavage, Anthony
; TITLE OF INVENTION: ISOLATED DROSOPHILA PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING DROSOPHILA PROTEINS, AND USES
; FILE REFERENCE: C1000048
; CURRENT APPLICATION NUMBER: US/60/142,845
; CURRENT FILING DATE: 1999-07-08
; NUMBER OF SEQ ID NOS: 704
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 289
```

```
; MEDIUM TYPE: PCT/US97/19575
; COMPUTER:
; OPERATING SYSTEM:
; SOFTWARE:
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US97/19575
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/739,150
; FILING DATE: 28-OCT-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/759,739
; FILING DATE: 06-DEC-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/891,928
; FILING DATE: 14-JULY-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: GTN-001CP10PC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 51:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1239 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1...1239
PCT-US97-19575-51

Alignment Scores:
Pred. No.: 25.1 Length: 1239
Score: 57.00 Matches: 12
Percent Similarity: 69.57% Conservative: 4
Best Local Similarity: 52.17% Mismatches: 7
Query Match: 44.19% Indels: 0
DB: 1 Gaps: 0
```

```
US-09-847-539a-6_COPY_59_86 (1-28) x PCT-US97-19575-51 (1-1239)
Qy 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValllys 23
Db 145 ATACAAGCCCTACAGAGCAAAATTCAGCGCTTAGATTCTCAAGAAAAAGTCGTTAGCAAA 204
Qy 24 AlaAspAsn 26
Db 205 TGGGATAAC 213

RESULT 15
US-08-759-739-250
; Sequence 250, Application US/08759739
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS SMITH ET AL
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; TITLE OF INVENTION: RELATING TO HELICOBACTER PYLORI VACCINE COMPOSITIONS THERE
; NUMBER OF SEQUENCES: 608
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street, Suite 510
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
```

```
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER:
; OPERATING SYSTEM:
; SOFTWARE:
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/759,739
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/487,032
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/561,469
; FILING DATE: 17-NOV-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/630,405
; FILING DATE: 01-APRIL-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/660,742
; FILING DATE: 06-JUN-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/736,791
; FILING DATE: 25-OCT-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/739,150
; FILING DATE: 28-OCT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: GTN-001CP8
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 250:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1239 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1...1239
US-08-759-739-250

Alignment Scores:
Pred. No.: 25.1 Length: 1239
Score: 57.00 Matches: 12
Percent Similarity: 69.57% Conservative: 4
Best Local Similarity: 52.17% Mismatches: 7
Query Match: 44.19% Indels: 0
DB: 1 Gaps: 0

US-09-847-539a-6_COPY_59_86 (1-28) x US-08-759-739-250 (1-1239)
Qy 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValllys 23
Db 145 ATACAAGCCCTACAGAGCAAAATTCAGCGCTTAGATTCTCAAGAAAAAGTCGTTAGCAAA 204
Qy 24 AlaAspAsn 26
Db 205 TGGGATAAC 213

RESULT 16
US-08-993-002A-4235
; Sequence 4235, Application US/08993002A
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS SMITH et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; TITLE OF INVENTION: RELATING TO HELICOBACTER PYLORI FOR
```



```
; APPLICATION NUMBER: US/08/759,739
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/487,032
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/561,469
; FILING DATE: 17-NOV-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/630,405
; FILING DATE: 01-APRIL-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/660,742
; FILING DATE: 06-JUN-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/736,791
; FILING DATE: 25-OCT-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/739,150
; FILING DATE: 28-OCT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: GTN-001CP8
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 92:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 666 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1...666
US-08-759-739-92
Alignment Scores:
Pred. No.: 12.3 Length: 666
Score: 57.00 Matches: 12
Percent Similarity: 69.57% Conservative: 4
Best Local Similarity: 52.17% Mismatches: 7
Query Match: 44.19% Indels: 0
DB: 11 Gaps: 0
US-08-847-539A-6_COPY_59_86 (1-28) x US-08-759-739-92 (1-666)
QY 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVallys 23
Db 229 ATCAAGCCCTACAGGAGCAAAATTCAGCGTTTAGATTCTCAAGAAAAAGTCGTTAGCAAA 288
QY 24 AlaAspAsn 26
Db 289 TGGGATAAC 297
RESULT 13
US-08-993-002A-4234
; Sequence 4234, Application US/08993002A
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS SMITH et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; TITLE OF INVENTION: RELATING TO HELICOBACTER PYLORI FOR
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 10031
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 State Street
```

```
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: Windows NT 4.0
; SOFTWARE: UNIX
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/993,002A
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: GTN-018
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)742-4214
; INFORMATION FOR SEQ ID NO: 4234:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 666 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1...666
US-08-993-002A-4234
Alignment Scores:
Pred. No.: 12.3 Length: 666
Score: 57.00 Matches: 12
Percent Similarity: 69.57% Conservative: 4
Best Local Similarity: 52.17% Mismatches: 7
Query Match: 44.19% Indels: 0
DB: 13 Gaps: 0
US-09-847-539A-6_COPY_59_86 (1-28) x US-08-993-002A-4234 (1-666)
QY 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVallys 23
Db 229 ATCAAGCCCTACAGGAGCAAAATTCAGCGTTTAGATTCTCAAGAAAAAGTCGTTAGCAAA 288
QY 24 AlaAspAsn 26
Db 289 TGGGATAAC 297
RESULT 14
PCT-US97-19575-51
; Sequence 51, Application PC/TUS9719575
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; TITLE OF INVENTION: RELATING TO HELICOBACTER PYLORI AND
; TITLE OF INVENTION: VACCINE COMPOSITIONS THEREOF
; NUMBER OF SEQUENCES: 208
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
```

```
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/561,469
; FILING DATE: 17-NOV-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/630,405
; FILING DATE: 01-APRIL-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/660,742
; FILING DATE: 06-JUN-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/736,791
; FILING DATE: 25-OCT-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/739,150
; FILING DATE: 28-OCT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: GTN-001CP8
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 43:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 576 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1...576
US-08-759-739-43
```

```
Alignment Scores:
Pred. No.: 10.4 Length: 576
Score: 57.00 Matches: 12
Percent Similarity: 69.57% Conservative: 4
Best Local Similarity: 52.17% Mismatches: 7
Query Match: 44.19% Indels: 0
DB: 11 Gaps: 0
```

```
US-09-847-539a-6_COPY_59_86 (1-28) x US-08-759-739-43 (1-576)
QY 4 LeuGluAlaLeuAlaaspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLys 23
Db 145 ATACAGCCCTACAGGCAAAATTGACGCTTTAGATTCTCAAGAAAAAAGTCGTTAGCAA 204
QY 24 AlaAspAsn 26
Db 205 TGGGATAAC 213
```

```
RESULT 11
US-08-993-002A-4232
; Sequence 4232, Application US/08993002A
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS SMITH et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; TITLE OF INVENTION: RELATING TO HELICOBACTER PYLORI FOR
; NUMBER OF SEQUENCES: 10031
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
```

```
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: Windows NT 4.0
; SOFTWARE: UNIX
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/993,002A
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: GTN-018
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)742-4214
; INFORMATION FOR SEQ ID NO: 4232:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 576 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1...576
US-08-993-002A-4232
```

```
Alignment Scores:
Pred. No.: 10.4 Length: 576
Score: 57.00 Matches: 12
Percent Similarity: 69.57% Conservative: 4
Best Local Similarity: 52.17% Mismatches: 7
Query Match: 44.19% Indels: 0
DB: 13 Gaps: 0
```

```
US-09-847-539a-6_COPY_59_86 (1-28) x US-08-993-002A-4232 (1-576)
QY 4 LeuGluAlaLeuAlaaspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLys 23
Db 145 ATACAGCCCTACAGGCAAAATTGACGCTTTAGATTCTCAAGAAAAAAGTCGTTAGCAA 204
QY 24 AlaAspAsn 26
Db 205 TGGGATAAC 213
```

```
RESULT 12
US-08-759-739-92
; Sequence 92, Application US/08759739
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS SMITH ET AL
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; TITLE OF INVENTION: RELATING TO HELICOBACTER PYLORI VACCINE COMPOSITIONS THERE
; NUMBER OF SEQUENCES: 608
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street, Suite 510
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER:
; OPERATING SYSTEM:
; SOFTWARE:
; CURRENT APPLICATION DATA:
```

```
; TITLE OF INVENTION: RELATING TO HELICOBACTER PYLORI FOR
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 880
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,032A
; FILING DATE: 07-JUNE-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: GTN-001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 282:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 576 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Helicobacter pylori
;
US-08-487-032A-282

Alignment Scores:
Pred. No.: 10.4 Length: 576
Score: 57.00 Matches: 12
Percent Similarity: 69.57% Conservative: 4
Best Local Similarity: 52.17% Mismatches: 7
Query Match: 44.19% Indels: 0
DB: 8 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x US-08-487-032A-282 (1-576)
Qy 4 LeuGluaLaLeuAlaSpGlnThrAspAlaLeuGlnSerGluGluAlaValVallys 23
Db 145 ATACAGCCCTACAGGCAAAATTCAGCTTTAGATTCTCAAGAAAAAGTCGTTAGCAA 204
Qy 24 AlaAspAsn 26
Db 205 TGGGATAAC 213

RESULT 9
US-08-561-469A-282
; Sequence 282, Application US/08561469A
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS SMITH
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; TITLE OF INVENTION: RELATING TO HELICOBACTER PYLORI FOR
; NUMBER OF SEQUENCES: 994
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street, Suite 510
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
```

```
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/561,469A
; FILING DATE: 17-NOV-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/487,032
; FILING DATE: 07-JUNE-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: GTN-001CP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 282:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 576 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Helicobacter pylori
;
US-08-561-469A-282

Alignment Scores:
Pred. No.: 10.4 Length: 576
Score: 57.00 Matches: 12
Percent Similarity: 69.57% Conservative: 4
Best Local Similarity: 52.17% Mismatches: 7
Query Match: 44.19% Indels: 0
DB: 9 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x US-08-561-469A-282 (1-576)
Qy 4 LeuGluaLaLeuAlaSpGlnThrAspAlaLeuGlnSerGluGluAlaValVallys 23
Db 145 ATACAGCCCTACAGGCAAAATTCAGCTTTAGATTCTCAAGAAAAAGTCGTTAGCAA 204
Qy 24 AlaAspAsn 26
Db 205 TGGGATAAC 213

RESULT 10
US-08-759-739-43
; Sequence 43, Application US/08759739
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS SMITH ET AL
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; TITLE OF INVENTION: RELATING TO HELICOBACTER PYLORI VACCINE COMPOSITIONS THERE
; NUMBER OF SEQUENCES: 608
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street, Suite 510
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER:
; OPERATING SYSTEM:
; SOFTWARE:
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/759,739
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/487,032
```

```

: APPLICANT: Rasmussen, Magnus
: TITLE OF INVENTION: STREPTOCOCCAL ALPHA 2M BINDING PROTEIN
: FILE REFERENCE: 100084.415US / N.75312B
: CURRENT APPLICATION NUMBER: US/09/847.539A
: CURRENT FILING DATE: 2001-05-01
: NUMBER OF SEQ ID NOS: 30
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 27
: LENGTH: 764
: TYPE: DNA
: ORGANISM: Streptococcus pyogenes
US-09-847-539A-27

```

Alignment Scores:	8.01e-12	Length:	764
Pred. No.:	Score:	Matches:	28
	129.00	Conservative:	0
Percent Similarity:	100.00%	Mismatches:	0
Best Local Similarity:	100.00%	Indels:	0
Query Match:	100.00%	Gaps:	0
DB:	32		

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-09-847-539A-27 (1-764)

[illegible]

## RESULT 5

```

US-09-847-539A-13
: Sequence 13, Application US/09847539A
: GENERAL INFORMATION:
: APPLICANT: Bjorck, Lars H
: APPLICANT: Rasmussen, Magnus
: TITLE OF INVENTION: STREPTOCOCCAL AL
: FILE REFERENCE: 100084.415US / N.753
: CURRENT APPLICATION NUMBER: US/09/84
: CURRENT FILING DATE: 2001-05-01
: NUMBER OF SEQ ID NOS: 30
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 13
: LENGTH: 777
: TYPE: DNA
: ORGANISM: Streptococcus pyogenes
US-09-847-539A-13

```

Alignment Scores:	8.17e-12	Length:	777
Pred. No.:	129.00	Matches:	28
Score:	100.00%	Conservative:	0
Percent Similarity:	100.00%	Mismatches:	0
Best Local Similarity:	100.00%	Indels:	0
Query Match:	100.00%	Gaps:	0
DB:	32		

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-09-847-539A-13 (1-777)

Qy	1	SerAspAlaLeuGluAlaLalaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaA	20
Db	319	AGTACGCCCTTAGAACGATTGGCGGATCAAAACAGACGCTTTACAATCAGAAGAAGCTGCG	378
Qy	21	ValValtysAlaAspAsnAlaA	28
Db	379	GTTGTGTTAAAGCGGATAACGCTGCT	402

## RESULT 6

US-09-847-539A-15  
; Sequence 15, Application US/09847539A  
; GENERAL INFORMATION:  
; APPLICANT: Bjorck, Lars H  
; APPLICANT: Rasmussen, Magnus

Alignment Scores:

Pred. No.:	9,09e-12	Length:	853
Score:	129.00%	Matches:	28
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	32	Gaps:	0

US-09-847-539A-6\_COPY\_59\_86 (1-28) X US-09-847-539A-15 (1-853)

Oy	1	SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaA	20
Db	556	AGTGACGCCCTTAGAAGCATTTGGCGGATCAAAACAGACGCTTTACAATCAGAAGAAGTGC	615
Oy	21	ValValLysAlaAspAsnAlaAlaA	28
Db	616	GTTGTGTTAAAGCGGATAACGCTGT	639

## RESULT 7

```

US-09-489-039A-318
; Sequence 318, Application US/09489039A
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND
; TITLE OF INVENTION: PNEUMONIAE FOR D
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 318
; LENGTH: 828
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-318

```

Alignment Scores:		
Pred. No.:	10.7	Length: 828
Score:	58.00	Matches: 12
Percent Similarity:	75.00%	Conservative: 6
Best Local Similarity:	50.00%	Mismatches: 6
Query Match:	44.96%	Indels: 0
DB:	19	Gaps: 0

US-09-847-539A-6 COPY 59 86 (1-28) x US-09-489-039A-318 (1-828)

Oy	2	AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaVal	21
Db	157	GACCGTCTGAGGCGCTCGCCGACAGCTCGAGCGCGCTACGCCATCCAGGCAAGCATC	216
Oy	22	VallysAlaAsp	25
Db	217	CTCAAGCCGCAT	228

PRECITY.T.B.

RESULTS 8  
US-08-487-032A-282  
; Sequence 282, Application US/08487032A  
; GENERAL INFORMATION:  
; APPLICANT: DOUGLAS SMITH  
; TITLE OF INVENTION: NUCLEIC ACID A

```
17 57 44.2 3497 10 US-08-621-425-49 Sequence 49, Appl
18 54 41.9 1783 53 US-60-142-845-289 Sequence 289, App
19 54 41.9 1785 53 US-60-145-134-143 Sequence 143, App
20 54 41.9 3046 56 US-60-173-464-14401 Sequence 14401, A
21 54 41.9 3049 55 US-60-167-217-17590 Sequence 17590, A
22 54 41.9 3349 23 US-09-614-150-17531 Sequence 17531, A
23 54 41.9 3349 58 US-60-191-637-17581 Sequence 17581, A
24 54 41.9 3349 58 US-60-191-681-13897 Sequence 13897, A
25 54 41.9 6398 23 US-09-614-150-17530 Sequence 17530, A
26 54 41.9 6398 58 US-60-191-637-17580 Sequence 17580, A
27 54 41.9 6398 58 US-60-191-681-13896 Sequence 13896, A
28 54 41.9 7203 55 US-60-173-464-14400 Sequence 14400, A
29 54 41.9 7206 55 US-60-167-217-17589 Sequence 17589, A
30 54 32275 19 US-09-528-237A-1811 Sequence 1811, Ap
31 52.5 40.7 1239 16 US-09-252-991A-8951 Sequence 8951, Ap
32 52.5 40.7 2127 16 US-09-252-991A-8732 Sequence 8732, Ap
33 52.5 40.7 8830 47 US-60-082-302-513 Sequence 513, App
34 52 40.3 594 22 US-09-565-309A-32821 Sequence 32821, A
35 52 40.3 641 22 US-09-565-309A-32824 Sequence 32824, A
36 52 40.3 939 22 US-09-565-309A-67148 Sequence 67148, A
37 52 40.3 984 28 US-09-708-427-29368 Sequence 29368, A
38 52 40.3 999 22 US-09-565-309A-67147 Sequence 67147, A
39 52 40.3 1230 22 US-09-565-309A-62025 Sequence 62025, A
40 52 40.3 1230 22 US-09-585-329A-545 Sequence 545, App
41 52 40.3 1244 22 US-09-565-309A-52656 Sequence 52656, A
42 52 40.3 16135 23 US-09-614-150-35057 Sequence 35057, A
43 52 40.3 16135 23 US-09-619-049-1340 Sequence 1340, Ap
44 52 40.3 16135 56 US-60-171-627-1946 Sequence 1946, Ap
45 52 40.3 16135 56 US-60-173-464-26803 Sequence 26803, A
```

## ALIGNMENTS

```
RESULT 1
US-09-847-539A-14
; Sequence 14, Application US/09847539A
; GENERAL INFORMATION:
; APPLICANT: Bjorck, Lars H
; APPLICANT: Rasmussen, Magnus
; TITLE OF INVENTION: STREPTOCOCCAL ALPHA ZM BINDING PROTEIN
; FILE REFERENCE: 100084.415US / N.75312B
; CURRENT APPLICATION NUMBER: US/09/847,539A
; CURRENT FILING DATE: 2001-05-01
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
; LENGTH: 469
; TYPE: DNA
; ORGANISM: Streptococcus pyogenes
US-09-847-539A-14
```

```
Alignment Scores:
Pred. No.: 4,57e-12 Length: 469
Score: 129.00 Matches: 28
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 32 Gaps: 0
```

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-09-847-539A-14 (1-469)

```
QY 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaLa 20
|||||
Db 172 TCAGATGCCTTAGAGCATTAGCGGATCAACAGACGCTTTACAATCACAAGAAGCTGCG 231
QY 21 ValValLysAlaAspAsnAlaLa 28
|||||
Db 232 GTTGTTAAAGCGGATAACGCTGCT 255
```

```
RESULT 2
US-09-847-539A-16
; Sequence 16, Application US/09847539A
; GENERAL INFORMATION:
```

```
; APPLICANT: Bjorck, Lars H
; APPLICANT: Rasmussen, Magnus
; TITLE OF INVENTION: STREPTOCOCCAL ALPHA ZM BINDING PROTEIN
; FILE REFERENCE: 100084.415US / N.75312B
; CURRENT APPLICATION NUMBER: US/09/847,539A
; CURRENT FILING DATE: 2001-05-01
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 16
; LENGTH: 504
; TYPE: DNA
; ORGANISM: Streptococcus pyogenes
US-09-847-539A-16
```

```
Alignment Scores:
Pred. No.: 4,96e-12 Length: 504
Score: 129.00 Matches: 28
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 32 Gaps: 0
```

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-09-847-539A-16 (1-504)

```
QY 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaLa 20
|||||
Db 208 TCAGATGCCTTAGAGCATTAGCGGATCAACAGACGCTTTACAATCACAAGAAGCTGCG 267
QY 21 ValValLysAlaAspAsnAlaLa 28
|||||
Db 268 GTTGTTAAAGCGGATAACGCTGCT 291
```

```
RESULT 3
US-09-847-539A-12
; Sequence 12, Application US/09847539A
; GENERAL INFORMATION:
; APPLICANT: Bjorck, Lars H
; APPLICANT: Rasmussen, Magnus
; TITLE OF INVENTION: STREPTOCOCCAL ALPHA ZM BINDING PROTEIN
; FILE REFERENCE: 100084.415US / N.75312B
; CURRENT APPLICATION NUMBER: US/09/847,539A
; CURRENT FILING DATE: 2001-05-01
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 12
; LENGTH: 654
; TYPE: DNA
; ORGANISM: Streptococcus pyogenes
US-09-847-539A-12
```

```
Alignment Scores:
Pred. No.: 6,7e-12 Length: 654
Score: 129.00 Matches: 28
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 32 Gaps: 0
```

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-09-847-539A-12 (1-654)

```
QY 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaLa 20
|||||
Db 274 TCAGATGCCTTAGAGCATTAGCGGATCAACAGACGCTTTACAATCACAAGAAGCTGCG 333
QY 21 ValValLysAlaAspAsnAlaLa 28
|||||
Db 334 GTTGTTAAAGCGGATAACGCTGCT 357
```

```
RESULT 4
US-09-847-539A-27
; Sequence 27, Application US/09847539A
; GENERAL INFORMATION:
; APPLICANT: Bjorck, Lars H
```

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 13, 2002, 04:57:28 ; Search time 558.053 Seconds  
(without alignments)  
1085.518 Million cell updates/sec

Title: US-09-847-539A-6\_COPY\_59\_86

Perfect score: 129

Sequence: 1 SDALEALADOTDQLQSEEAIVKADNAA 28

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 21979536 seqs, 10817449327 residues

Total number of hits satisfying chosen parameters: 43959072

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool/US09847539/runat\_10102002\_093107\_5059/app\_query.fasta\_1.526  
-DB=Pending Patents\_NA\_Main -QPMT=fastap -SUFFIX=std.rnmp -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosu62  
-TRANS=human40.cdd -LIST=45 -DOALIGN=200 -THR\_SCORE=pct.THR\_MAX=100  
-THR\_MIN=0 -ALIGN=40 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=US09847539.ecgn\_1\_1201.brnat\_10102002\_093107\_5059  
-NCPU=6 -ICPU=3 -NO\_XLPY -NO\_MMAL -LARGQUERY -NEG\_SCORES=0 -WAIT -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: /cgn2\_6/ptodata/2/pna/PCTUS\_COMB.seq.\*  
2: /cgn2\_6/ptodata/2/pna/US06\_COMB.seq.\*  
3: /cgn2\_6/ptodata/2/pna/US07\_COMB.seq.\*  
4: /cgn2\_6/ptodata/2/pna/US080\_COMB.seq.\*  
5: /cgn2\_6/ptodata/2/pna/US081\_COMB.seq.\*  
6: /cgn2\_6/ptodata/2/pna/US082\_COMB.seq.\*  
7: /cgn2\_6/ptodata/2/pna/US083\_COMB.seq.\*  
8: /cgn2\_6/ptodata/2/pna/US084\_COMB.seq.\*  
9: /cgn2\_6/ptodata/2/pna/US085\_COMB.seq.\*  
10: /cgn2\_6/ptodata/2/pna/US086\_COMB.seq.\*  
11: /cgn2\_6/ptodata/2/pna/US087\_COMB.seq.\*  
12: /cgn2\_6/ptodata/2/pna/US088\_COMB.seq.\*  
13: /cgn2\_6/ptodata/2/pna/US089\_COMB.seq.\*  
14: /cgn2\_6/ptodata/2/pna/US090\_COMB.seq.\*  
15: /cgn2\_6/ptodata/2/pna/US091\_COMB.seq.\*  
16: /cgn2\_6/ptodata/2/pna/US092\_COMB.seq.\*  
17: /cgn2\_6/ptodata/2/pna/US093\_COMB.seq.\*  
18: /cgn2\_6/ptodata/2/pna/US094\_COMB.seq.\*  
19: /cgn2\_6/ptodata/2/pna/US095A\_COMB.seq.\*  
20: /cgn2\_6/ptodata/2/pna/US095B\_COMB.seq.\*  
21: /cgn2\_6/ptodata/2/pna/US095C\_COMB.seq.\*  
22: /cgn2\_6/ptodata/2/pna/US095D\_COMB.seq.\*  
23: /cgn2\_6/ptodata/2/pna/US096A\_COMB.seq.\*  
24: /cgn2\_6/ptodata/2/pna/US096B\_COMB.seq.\*  
25: /cgn2\_6/ptodata/2/pna/US096C\_COMB.seq.\*  
26: /cgn2\_6/ptodata/2/pna/US096D\_COMB.seq.\*  
27: /cgn2\_6/ptodata/2/pna/US096E\_COMB.seq.\*  
28: /cgn2\_6/ptodata/2/pna/US097A\_COMB.seq.\*

29: /cgn2\_6/ptodata/2/pna/US097B\_COMB.seq.\*  
30: /cgn2\_6/ptodata/2/pna/US097C\_COMB.seq.\*  
31: /cgn2\_6/ptodata/2/pna/US098A\_COMB.seq.\*  
32: /cgn2\_6/ptodata/2/pna/US098B\_COMB.seq.\*  
33: /cgn2\_6/ptodata/2/pna/US098C\_COMB.seq.\*  
34: /cgn2\_6/ptodata/2/pna/US099A\_COMB.seq.\*  
35: /cgn2\_6/ptodata/2/pna/US099B\_COMB.seq.\*  
36: /cgn2\_6/ptodata/2/pna/US099C\_COMB.seq.\*  
37: /cgn2\_6/ptodata/2/pna/US100\_COMB.seq.\*  
38: /cgn2\_6/ptodata/2/pna/US101\_COMB.seq.\*  
39: /cgn2\_6/ptodata/2/pna/US6000\_COMB.seq.\*  
40: /cgn2\_6/ptodata/2/pna/US6001\_COMB.seq.\*  
41: /cgn2\_6/ptodata/2/pna/US6002\_COMB.seq.\*  
42: /cgn2\_6/ptodata/2/pna/US6003\_COMB.seq.\*  
43: /cgn2\_6/ptodata/2/pna/US6004\_COMB.seq.\*  
44: /cgn2\_6/ptodata/2/pna/US6005\_COMB.seq.\*  
45: /cgn2\_6/ptodata/2/pna/US6006\_COMB.seq.\*  
46: /cgn2\_6/ptodata/2/pna/US6007\_COMB.seq.\*  
47: /cgn2\_6/ptodata/2/pna/US6008\_COMB.seq.\*  
48: /cgn2\_6/ptodata/2/pna/US6009\_COMB.seq.\*  
49: /cgn2\_6/ptodata/2/pna/US6010\_COMB.seq.\*  
50: /cgn2\_6/ptodata/2/pna/US6011\_COMB.seq.\*  
51: /cgn2\_6/ptodata/2/pna/US6012\_COMB.seq.\*  
52: /cgn2\_6/ptodata/2/pna/US6013\_COMB.seq.\*  
53: /cgn2\_6/ptodata/2/pna/US6014\_COMB.seq.\*  
54: /cgn2\_6/ptodata/2/pna/US6015\_COMB.seq.\*  
55: /cgn2\_6/ptodata/2/pna/US6016\_COMB.seq.\*  
56: /cgn2\_6/ptodata/2/pna/US6017\_COMB.seq.\*  
57: /cgn2\_6/ptodata/2/pna/US6018\_COMB.seq.\*  
58: /cgn2\_6/ptodata/2/pna/US6019\_COMB.seq.\*  
59: /cgn2\_6/ptodata/2/pna/US6020\_COMB.seq.\*  
60: /cgn2\_6/ptodata/2/pna/US6021\_COMB.seq.\*  
61: /cgn2\_6/ptodata/2/pna/US6022\_COMB.seq.\*  
62: /cgn2\_6/ptodata/2/pna/US6023\_COMB.seq.\*  
63: /cgn2\_6/ptodata/2/pna/US6024\_COMB.seq.\*  
64: /cgn2\_6/ptodata/2/pna/US6025\_COMB.seq.\*  
65: /cgn2\_6/ptodata/2/pna/US6026\_COMB.seq.\*  
66: /cgn2\_6/ptodata/2/pna/US6027\_COMB.seq.\*  
67: /cgn2\_6/ptodata/2/pna/US6028\_COMB.seq.\*  
68: /cgn2\_6/ptodata/2/pna/US6029\_COMB.seq.\*  
69: /cgn2\_6/ptodata/2/pna/US6030\_COMB.seq.\*  
70: /cgn2\_6/ptodata/2/pna/US6031\_COMB.seq.\*  
71: /cgn2\_6/ptodata/2/pna/US6032\_COMB.seq.\*  
72: /cgn2\_6/ptodata/2/pna/US6033\_COMB.seq.\*  
73: /cgn2\_6/ptodata/2/pna/US6034\_COMB.seq.\*  
74: /cgn2\_6/ptodata/2/pna/US6035\_COMB.seq.\*  
75: /cgn2\_6/ptodata/2/pna/US6036\_COMB.seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	129	100.0	469	32	US-09-847-539A-14
2	129	100.0	504	32	US-09-847-539A-16
3	129	100.0	654	32	US-09-847-539A-12
4	129	100.0	764	32	US-09-847-539A-27
5	129	100.0	777	32	US-09-847-539A-13
6	129	100.0	853	32	US-09-847-539A-15
7	58	45.0	828	18	US-09-489-039A-318
8	57	44.2	576	9	US-08-487-032A-282
9	57	44.2	576	9	US-08-561-469A-282
10	57	44.2	576	11	US-08-759-739-43
11	57	44.2	576	13	US-08-993-002A-4332
12	57	44.2	666	11	US-08-759-739-92
13	57	44.2	666	13	US-08-993-002A-4234
14	57	44.2	1239	1	PCT-US97-19575-51
15	57	44.2	1239	11	US-08-759-739-250
16	57	44.2	1239	13	US-08-993-002A-4235

```
Db 1594 GAGCAGAAACTTCAGGATCTTCAAGGAGAAAGGATGCTTGGATTCTGAAAAAGCAGCAG 1653
Qy 107 ValValGlnSerAspAsnAlaAlaSerAspAlaTrpGluLysAlaAlaThrProIleAla 126
Db 1654 ATCACTGCACAGAAACAAGACCTGGAGGCAGAGGTGTCCAAGCTGCACAGAGAGAGGTTGCC 1713
Qy 127 LeuAspValLysLysThrLysAspThrLys 136
Db 1714 AAGCTGTCAAAAGAACTAGAAGATGCCAAG 1743
```

Search completed: October 13, 2002, 03:34:51  
Job time : 82.8663 secs

```

1  APPLICANT: King, Mary-Claire
2  APPLICANT: Lynch, Eric D.
3  APPLICANT: Lee, Ming
4  APPLICANT: Morrow, Jan E.
5  APPLICANT: Welsh, Piri L.
6  APPLICANT: Leon, Pedro E.
7  TITLE OF INVENTION: Modulators of Actin
8  NUMBER OF SEQUENCES: 14
9  CORRESPONDENCE ADDRESS:
10 ADDRESS: SCIENCE & TECHNOLOGY LAW GROUP
11 STREET: 75 DENISE DRIVE
12 CITY: HILLSBOROUGH
13 STATE: CALIFORNIA
14 COUNTRY: USA
15 ZIP: 94010
16 COMPUTER READABLE FORM:
17 MEDIUM TYPE: Floppy disk
18 COMPUTER: IBM PC compatible
19 OPERATING SYSTEM: PC-DOS/MS-DOS
20 SOFTWARE: PatentIn Release #1.0, Version #1.30
21 CURRENT APPLICATION DATA:
22 APPLICATION NUMBER: US/09/323,735
23 FILING DATE:
24 CLASSIFICATION:
25 PRIOR APPLICATION DATA:
26 APPLICATION NUMBER: 09/080,897
27 FILING DATE:
28 ATTORNEY/AGENT INFORMATION:
29 NAME: OSMAN, RICHARD A.
30 REGISTRATION NUMBER: 36,627
31 REFERENCE/DOCKET NUMBER: UW97-001
32 TELECOMMUNICATION INFORMATION:
33 TELEPHONE: (650) 343-4341
34 TELEFAX: (650) 343-4342
35 INFORMATION FOR SEQ ID NO: 3:
36 SEQUENCE CHARACTERISTICS:
37 LENGTH: 4378 base pairs
38 TYPE: nucleic acid
39 STRANDEDNESS: double
40 TOPOLOGY: linear
41 MOLECULE TYPE: cdna
42 US-09-323-735-3

```

```

RESULT 40
US-08-899-595-2
; Sequence 2, Application US/08899595
; Patent No. 6111072
; GENERAL INFORMATION:
; APPLICANT: Narumiya, Shuh
; APPLICANT: Takahashi, No. 6111072uakl
; TITLE OF INVENTION: RHO TARGET PROTEIN HUMAN MDIA AND GENE
; TITLE OF INVENTION: ENCODING SAME
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/899,595
; FILING DATE: 24-JUL-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 8-242701
; FILING DATE: 26-AUG-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 9-90170
; FILING DATE: 25-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Stephen A. Bent
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 049441/0112
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4399 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; ORIGINAL SOURCE:
; ORGANISM: Mouse
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 94..3858
US-08-899-595-2

Alignment Scores:
Pred. No.:          39.8      Length:       4399
Score:             62.00     Matches:        20
Percent Similarity: 40.00%    Conservative:   16
Best Local Similarity: 22.22% Mismatches:       54
Query Match:       7.98%     Indels:         0
DB:                3        Gaps:           0

US-09-847-539A-6 (1-159) x US-08-899-595-2 (1-4399)

Qy  47 AspLysGluAlaThrThrAlaIleGluAlaLaSerSerAspAlaLeuGluAla
|||||              ||||| : : : : : : : : : : : :
Db 1474 GATAAACAAAGGTGGAAAAATCTGAGGCCAAAGCTACAGACTGGAAAAAAGT
                                     |
Qy  67 AspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValAlaAlaAspAsn
||||| : : : : : : : : : : : :
Db 1534 TCAGAATTAAACAGCGCCGCACGAGTTACAGTAGAATAATGAANAAGATGGAATAAT
                                     |
Qy  87 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGlu

```



APPLICANT: Hackensack  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07601  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/678,614  
FILING DATE: 10-JUL-1996  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Jackson Esq., David A.  
REGISTRATION NUMBER: 26,742  
REFERENCE/DOCKET NUMBER: 600-1-139  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-487-5800  
TELEFAX: 201-343-1684  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4248 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Staphylococcus aureus  
STRAIN: RUSA 266  
FEATURE:  
NAME/KEY: ORF  
LOCATION: 154..1410  
FEATURE:  
NAME/KEY: ORF  
LOCATION: 1497..3500  
US-08-678-614-1

Alignment Scores:  
Pred. No.: 38 Length: 4248  
Score: 62.00 Matches: 18  
Percent Similarity: 41.56% Conservative: 14  
Best Local Similarity: 23.38% Mismatches: 45  
Query Match: 7.98% Indels: 0  
DB: 3 Gaps: 0

US-09-847-539A-6 (1-159) x US-08-678-614-1 (1-4248)

Qy 45 IleGluAspLysGluAlaThrAlaIleGluAlaSerSerAspAlaLeuGluAla 64  
Db 2343 ATTGATGCAGCAACAAATGCACAGTAGACCCATTAACAAAGCAATCAATGAT 2402  
Qy 65 LeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaValValLysAlaAspAsn 84  
Db 2403 ATTAATCAAACTACACCTGCTACACAGCTTAACAGCAGCTCTTGAAGAATTTGACGAA 2462  
Qy 85 AlaAlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGlu 104  
Db 2463 GTTGTTCAGCACAATATGATCAAGACCTTTAAATCCTGATACACAAATGAAGAAGTA 2522  
Qy 105 AlaGluValValGlnSerAspAsnAlaSerAspAlaTrpGluLysAla 121  
Db 2523 GCGAAGCTATTGAACGTATTATTAATGCAGCTAAAGTTCTGCTGTTAAAGCA 2573

RESULT 38  
US-09-080-897-3  
Sequence 3, Application US/09080897  
Patent No. 5985574  
GENERAL INFORMATION:  
APPLICANT: King, Mary-Claire  
Lynch, Eric D.

APPLICANT: Lee, Ming  
APPLICANT: Morrow, Jan E.  
APPLICANT: Welsh, Piri L.  
APPLICANT: Leon, Pedro E.  
TITLE OF INVENTION: Modulators of Actin  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP  
STREET: 75 DENISE DRIVE  
CITY: HILLSBOROUGH  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 94010  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/080,897  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: OSMAN, RICHARD A  
REGISTRATION NUMBER: 36,627  
REFERENCE/DOCKET NUMBER: UW97-001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (650) 343-4341  
TELEFAX: (650) 343-4342  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4378 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
US-09-080-897-3

Alignment Scores:  
Pred. No.: 39.6 Length: 4378  
Score: 62.00 Matches: 20  
Percent Similarity: 40.00% Conservative: 16  
Best Local Similarity: 22.22% Mismatches: 54  
Query Match: 7.98% Indels: 0  
DB: 2 Gaps: 0

US-09-847-539A-6 (1-159) x US-09-080-897-3 (1-4378)

Qy 47 AspLysGluAlaThrAlaIleGluAlaSerSerAspAlaLeuGluAlaLeuAla 66  
Db 1474 GATAAACAAGGTGGGAAAAATCTGAGCCCAAGCTACAGAGCTGGAAAAAGTTGGAT 1533  
Qy 67 AspGlnThrAspAlaLeuGlnSerGluGluAlaValValLysAlaAspAlaAla 86  
Db 1534 TCAGAAATTAACGCGCGCACAGAGTTTACAAGTAGAAATGAAAGAGTGGAAATGACTTT 1593  
Qy 87 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaGlu 106  
Db 1594 GAGCAGAACTTCAGGATCTTCAAGGAGAAAGAGTGCCTTGGATTCTGAAAACGACGAG 1653  
Qy 107 ValValGlnSerAspAsnAlaAlaSerAspAlaTrpGluLysAlaAlaThrProIleAla 126  
Db 1654 ATCACTGCACAGAAACAAGACCTGGAGGACAGGTGTCCAAGCTGCACAGGAGAGTTGCC 1713  
Qy 127 LeuAspValLysLysThrLysAspThrLys 136  
Db 1714 AAGCTGTCAAAAGAACTAGAAAGATGCCAAG 1743

RESULT 39  
US-09-323-735-3  
Sequence 3, Application US/09323735  
Patent No. 6197932  
GENERAL INFORMATION:  
APPLICANT: Lynch, Eric D.









```
;
;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5200 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 118..183
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 118..3699
; FEATURE:
; NAME/KEY: polyA_site
; LOCATION: 4433
; FEATURE:
; NAME/KEY: polyA_site
; LOCATION: 5195
;
; US-08-317-450B-12
;
Alignment Scores:
Pred. No.: 25.1 Length: 5200
Score: 64.00 Matches: 20
Percent Similarity: 45.16% Conservative: 8
Best Local Similarity: 32.26% Mismatches: 34
Query Match: 8.24% Indels: 0
DB: 1 Gaps: 0

US-09-847-539A-6 (1-159) x US-08-317-450B-12 (1-5200)
QY 30 LeuArgAsnGluGluArgAlaIleAspGluLeuLysLysGlnAlaIleGluAspLysGlu 49
Db 2980 CTCAGAGAGTTTGACCTGCAGGTGGACACACAGAAAGCAGAGCTCAAGAAGCCATGAAG 3039
QY 50 AlaThrThraIleGluAlaLaserSerAspAlaLeuGluAlaLeuAlaAspGlnThr 69
Db 3040 AGACTCTCTACATCAGCCAGCAAGAGTTTCAGATGCCAGTGACAAAGACCCAGCAAGCAGAA 3099
QY 70 AspAlaLeuGlnSerGluGluAlaAlaValValLysAlaAspAsnAlaLaserAspAla 89
Db 3100 AGAGCCCTGGGGAGCGCTCTGCTGATGCACAGAGGGCAAGAAGTGGCGCGGGAGGCC 3159
QY 90 LeuGlu 91
Db 3160 CTGGAA 3165

RESULT 27
US-08-800-593-12
; Sequence 12, Application US/08800593
; Patent No. 6143505
; GENERAL INFORMATION:
; APPLICANT: Tryggvason, Karl
; APPLICANT: Kallunki, Pekka
; APPLICANT: Pyke, Charles
; TITLE OF INVENTION: Laminin Chalos: Diagnostic and
; TITLE OF INVENTION: Therapeutic Use
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
; STREET: 300 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/800,593
; FILING DATE: 18-FEB-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/317,450
; FILING DATE: 04-OCT-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Chao, Mark
; REGISTRATION NUMBER: 37,293
; REFERENCE/DOCKET NUMBER: 94,778-B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-913-0001
; TELEFAX: 312-913-0002
; INFORMATION FOR SEQ ID NO: 12:
;
;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5200 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 118..183
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 118..3699
; FEATURE:
; NAME/KEY: polyA_site
; LOCATION: 4433
; FEATURE:
; NAME/KEY: polyA_site
; LOCATION: 5195
;
; US-08-800-593-12
;
Alignment Scores:
Pred. No.: 25.1 Length: 5200
Score: 64.00 Matches: 20
Percent Similarity: 45.16% Conservative: 8
Best Local Similarity: 32.26% Mismatches: 34
Query Match: 8.24% Indels: 0
DB: 1 Gaps: 0

US-09-847-539A-6 (1-159) x US-08-800-593-12 (1-5200)
QY 30 LeuArgAsnGluGluArgAlaIleAspGluLeuLysLysGlnAlaIleGluAspLysGlu 49
Db 2980 CTCAGAGAGTTTGACCTGCAGGTGGACACACAGAAAGCAGAGCTCAAGAAGCCATGAAG 3039
QY 50 AlaThrThraIleGluAlaLaserSerAspAlaLeuGluAlaLeuAlaAspGlnThr 69
Db 3040 AGACTCTCTACATCAGCCAGCAAGAGTTTCAGATGCCAGTGACAAAGACCCAGCAAGCAGAA 3099
QY 70 AspAlaLeuGlnSerGluGluAlaAlaValValLysAlaAspAsnAlaLaserAspAla 89
Db 3100 AGAGCCCTGGGGAGCGCTCTGCTGATGCACAGAGGGCAAGAAGTGGCGCGGGAGGCC 3159
QY 90 LeuGlu 91
Db 3160 CTGGAA 3165

RESULT 28
US-08-961-858-2
; Sequence 2, Application US/08961858
; Patent No. 5834210
; GENERAL INFORMATION:
; APPLICANT: Liu, Shigui
; APPLICANT: Shi, Qiuwei
; TITLE OF INVENTION: STABLE TROPONIN SUBUNITS AND COMPLEXES
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,858
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
```

Qy 50 AlathrThrAlaIleGluAlaAlaSerSerAspAlaLeuGluAlaLeuAlaAspGlnThr 69  
Db 3040 AGACTCTCTACAGCCAGCAGAGGTTTCAGATGCCAGTGCACAGACCCACACAGCAGAA 3099  
Qy 70 AspAlaLeuGlnSerGluGluAlaAlaValVallyAlaAspAsnAlaAlaSerAspAla 89  
Db 3100 AGAGCCCTGGGAGGCGTCTGCTGATGCACAGAGGCAAGAAATGGGGCCGGGAGGCC 3159  
Qy 90 LeuGlu 91  
Db 3160 CTGGAA 3165  
RESULT 25  
US-08-800-593-14  
; Sequence 14, Application US/08800593  
; Patent No. 6143505  
; GENERAL INFORMATION:  
; APPLICANT: Tryggvason, Karl  
; APPLICANT: Kallunki, Pekka  
; APPLICANT: Pyke, Charles  
; TITLE OF INVENTION: Laminin Chains: Diagnostic and  
; TITLE OF INVENTION: Therapeutic Use  
; NUMBER OF SEQUENCES: 19  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff  
; STREET: 300 South Wacker Drive  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: USA  
; ZIP: 60606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/800,593  
; FILING DATE: 18-FEB-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/317,450  
; FILING DATE: 04-OCT-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Chao, Mark  
; REGISTRATION NUMBER: 37,293  
; REFERENCE/DOCKET NUMBER: 94,778-B  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312-913-0001  
; TELEFAX: 312-913-0002  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4316 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: CDNA  
; FEATURE:  
; NAME/KEY: sig\_peptide  
; LOCATION: 118..183  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 118..3453  
; FEATURE:  
; NAME/KEY: repeat\_unit  
; LOCATION: 4021..4316  
; OTHER INFORMATION: /rpt\_type="other"  
; OTHER INFORMATION: /rpt\_family="HUMAN ALU"  
; FEATURE:  
; NAME/KEY: polyA\_site  
; LOCATION: 4296  
US-08-800-593-14

Alignment Scores:  
Pred. No.: 19.6 Length: 4316  
Score: 64.00 Matches: 20  
Percent Similarity: 45.16% Conservative: 8  
Best Local Similarity: 32.26% Mismatches: 34  
Query Match: 8.24% Indels: 0  
DB: 3 Gaps: 0  
US-09-847-539a-6 (1-159) x US-08-800-593-14 (1-4316)  
Qy 30 LeuArgasngluGluArgAlaIleAspGluLeuLysLysGlnAlaIleGluAspLysGlu 49  
Db 2980 CTCAGAGAGTTTTCAGCTGCAGTGCACACAGAAAGCAGAGCTGAAGAAGCCATGAAG 3039  
Qy 50 AlathrThrAlaIleGluAlaAlaSerSerAspAlaLeuGluAlaLeuAlaAspGlnThr 69  
Db 3040 AGACTCTCTACATCAGCCAGCAGGTTTCAGATGCCAGTGCACAGACCCACACAGCAGAA 3099  
Qy 70 AspAlaLeuGlnSerGluGluAlaAlaValVallyAlaAspAsnAlaAlaSerAspAla 89  
Db 3100 AGAGCCCTGGGAGGCGTCTGCTGATGCACAGAGGCAAGAAATGGGGCCGGGAGGCC 3159  
Qy 90 LeuGlu 91  
Db 3160 CTGGAA 3165  
RESULT 26  
US-08-317-450B-12  
; Sequence 12, Application US/08317450B  
; Patent No. 5660982  
; GENERAL INFORMATION:  
; APPLICANT: Tryggvason, Karl  
; APPLICANT: Kallunki, Pekka  
; APPLICANT: Pyke, Charles  
; TITLE OF INVENTION: Laminin Chains: Diagnostic and  
; TITLE OF INVENTION: Therapeutic Use  
; NUMBER OF SEQUENCES: 19  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BANNER & ALLEGRETTI, LTD.  
; STREET: Ten South Wacker Drive  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: USA  
; ZIP: 60606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/317,450B  
; FILING DATE: 04-OCT-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Chao, Mark  
; REGISTRATION NUMBER: 37,293  
; REFERENCE/DOCKET NUMBER: 94,778  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312-715-1000  
; TELEFAX: 312-715-1234  
; INFORMATION FOR SEQ ID NO: 12:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 5200 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: CDNA  
; FEATURE:  
; NAME/KEY: sig\_peptide  
; LOCATION: 118..183  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 118..3699







```
Db 19431 CAGGAGCAGGAAGACAGCAATTAGAGGAGGTGGAGGACGAAGACGAGCAGGAGGAG 19372
Qy 130 LysLysThrLysAspThrLysProValValLysLysGluGluArgGln 145
Db 19371 CAGGAGGAGCAGGAGTTAGAGGAGGTGGAAGACGAGGAAGACGAGGAG 19324
RESULT 19
US-09-230-371A-20/c
; Sequence 20, Application US/09230371A
; Patent No. 6348586
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Bohenzky, Roy A
; APPLICANT: Russo, James J
; APPLICANT: Edelman, Isidore S
; APPLICANT: Moore, Patrick S
; TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS SEQUENCES AND
; FILE REFERENCE: 45185-G-PCT-US
; CURRENT APPLICATION NUMBER: US/09/230,371A
; CURRENT FILING DATE: 1999-11-17
; PRIOR APPLICATION NUMBER: PCT/US97/13346
; PRIOR FILING DATE: 1997-07-22
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 20
; LENGTH: 32207
; TYPE: DNA
; ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-09-230-371A-20
Alignment Scores:
Pred. No.: 206 Length: 32207
Score: 65.00 Matches: 18
Percent Similarity: 45.69% Conservative: 35
Best Local Similarity: 15.52% Mismatches: 63
Query Match: 8.37% Indels: 0
DB: 4 Gaps: 0
US-09-847-539A-6 (1-159) x US-09-230-371A-20 (1-32207)
Qy 30 LeuArgAsnGluGluArgAlaIleAspGluLeuLysGlnAlaIleGluAspLysGlu 49
Db 19671 TTAGAGGATCAGGACGAGGTAGAGGAGCAGGAGGAGGTAGAGGAGCAGGAGGAG 19612
Qy 50 AlaThrThrAlaIleGluAlaSerSerAspAlaLeuGluAlaLeuAlaAspGlnThr 69
Db 19611 CAGTTAGAGGAGCAGGAGGAGGTAGAGGAGCAGGAGGAGGTAGAGGAGCAGGAG 19552
Qy 70 AspAlaLeuGlnSerGluGluAlaValValLysAlaAspAsnAlaSerAspAla 89
Db 19551 CAGGAGTTAGAGGAGCAGGAGGAGGTAGAGGAGCAGGAGGAGGTAGAGGAGGAG 19492
Qy 90 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaGluValGln 109
Db 19491 GAGCAGGAGTTAGAGGAGCAGGAGGTGAGAGCAGGAGGAGGTGAGAGCAGGAGGAG 19432
Qy 110 SerAspAsnAlaAlaSerAspAlaTrpGluTysAlaAlaThrProIleAlaLeuAspVal 129
Db 19431 CAGGAGCAGGAAGACGAGCAATTAGAGGAGGTGGAGGAGCAAGACGAGGAGGAGGAG 19372
Qy 130 LysLysThrLysAspThrLysProValValLysLysGluGluArgGln 145
Db 19371 CAGGAGGAGCAGGAGTTAGAGGAGGTGGAAGACGAGGAAGACGAGGAG 19324
RESULT 20
US-08-216-894-1
; Sequence 1, Application US/08216894
; Patent No. 5876734
; GENERAL INFORMATION:
; APPLICANT: Kirchhoff, Louis V.
; APPLICANT: Otsu, Keiko
; TITLE OF INVENTION: POLYPEPTIDES FOR DIAGNOSING INFECTION
```

```
; TITLE OF INVENTION: WITH TRYPA NOSOMA CRUZI
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/216,894
; FILING DATE: 24-MAR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 85326/102/DRLO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1695 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1692
US-08-216-894-1
Alignment Scores:
Pred. No.: 5.58 Length: 1695
Score: 64.00 Matches: 19
Percent Similarity: 41.49% Conservative: 20
Best Local Similarity: 20.21% Mismatches: 55
Query Match: 8.24% Indels: 0
DB: 2 Gaps: 0
US-09-847-539A-6 (1-159) x US-08-216-894-1 (1-1695)
Qy 29 AlaLeuArgAsnGluGluArgAlaIleAspGluLeuLysLysGlnAlaIleGluAspLys 48
Db 1099 GCCTGGAGACCGGAGACGAGGAGGCGAGCTGAAGCCACGAGGTTGCCGACGGAAG 1158
Qy 49 GluAlaThrThrAlaIleGluAlaAlaSerSerAspAlaLeuGluAlaLeuAlaAspGln 68
Db 1159 CAGAAGGCGAGCTGAGCGCCGCAAGCGCTGGAGCAGGAGCAGGCGAGCTGAAGCC 1218
Qy 69 ThrAspAlaLeuGlnSerGluGluAlaAlaValValLysAlaAspAsnAlaAlaSerAsp 88
Db 1219 ACGAAGGTTGCCGACGCGGAGGAAGACGAGGAGGCGAGCTGAAGGCTGAAGGTTCC 1278
Qy 89 AlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaGluValVal 108
Db 1279 GAGAAGCAGAGGAGCTGAGGCCACGAGGAGGTTGCCGAAGCGGAGGAGCAGGAGCT 1338
Qy 109 GlnSerAspAsnAlaAlaSerAspAlaTrpGluLysAlaAla 122
Db 1339 GAAGCCACGAGGTTGCCGACGCGGAGGAAGACGAGGAGCAGCT 1380
RESULT 21
US-09-115-746-1
; Sequence 1, Application US/09115746
; Patent No. 6228601
; GENERAL INFORMATION:
; APPLICANT: Kirchhoff, Louis V.
; APPLICANT: Otsu, Keiko
```

```
;
; REGISTRATION NUMBER: 32,135
; REFERENCE/DOCKET NUMBER: P-3283
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15664 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 222..425
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 451..747
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 747..1109
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1109..2014
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2034..2747
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2747..3109
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 3109..3444
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 3444..3728
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 3731..4855
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 4855..5376
; OTHER INFORMATION: /function= "potential coding
; OTHER INFORMATION: sequence"
; OTHER INFORMATION: /product= "L5 gp37 homolog"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 5382..5747
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 5837..6307
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 6403..7770
```

```
;
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 7770..8006
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 8033..8236
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 8244..9443
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 9450..10244
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 10371..10586
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 11115..11786
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 11917..12741
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 12748..14499
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 14771..15154
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 15154..15426
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 15429..15664
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; US-08-402-282-3
;
; Alignment Scores:
; Pred. No.: 2.66e+03 Length: 15664
; Score: 42.00 Matches: 8
; Percent Similarity: 63.64% Conservative: 6
; Best Local Similarity: 36.36% Mismatches: 8
; Query Match: 32.56% Indels: 0
; Db: 1 Gaps: 0
;
; US-09-847-539A-6_COPY_59_86 (1-28) x US-08-402-282-3 (1-15664)
;
; Qy 2 AspAlaLeuGluAlaLeuAlaAspClnThrAspAlaLeuGlnSerGluGluAlaVal 21
;
; Db 5643 GAGGGCGCTGGCCACATCGCAGGACNAGCTCGACGTGTGTCAGACGACGCGCGCG 5702
```



; CITY: Chicago  
; STATE: IL  
; COUNTRY: US  
; ZIP: 60601-6780  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release. #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/581,148C  
; FILING DATE: 29-DEC-1995  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Larcher, Carol  
; REGISTRATION NUMBER: 35243  
; REFERENCE/DOCKET NUMBER: 71380  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (312) 616-5600  
; TELEFAX: (312) 616-5700  
; INFORMATION FOR SEQ ID NO: 15:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1903 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..1665  
; US-08-581-148C-15

Alignment Scores:  
Pred. No.: 195 Length: 1903  
Score: 42.00 Matches: 9  
Percent Similarity: 60.00% Conservative: 6  
Best Local Similarity: 36.00% Mismatches: 10  
Query Match: 32.56% Indels: 0  
DB: 3 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-08-581-148C-15 (1-1903)

QY 2 AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluAlaLaVal 21  
::: |||||::: ||| ||| :::  
Db 918 GAAATACCTGGAAGCCCTAGCGGGACGCTCCAGGCTGCTGGAGGAGCGCGGAGCTT 859  
QY 22 ValLysAlaAspAsn 26  
::: |||||::: ||| ||| :::  
Db 858 ATAGAAGCTGACAGT 844

RESULT 13  
US-08-285-641-20  
; Sequence 20, Application US/08285641  
; Patent No. 5726299  
; GENERAL INFORMATION:  
; APPLICANT: Zupancic, Thomas J.  
; APPLICANT: Yukawa, Hideaki  
; TITLE OF INVENTION: PROMOTER DNA FRAGMENT FROM CORNYNEFORM BACTERIA  
; NUMBER OF SEQUENCES: 24  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sughrue, Mion, Zinn, Macpeak & Seas  
; STREET: 2100 Pennsylvania Avenue, N.W.  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: United States  
; ZIP: 20037-3202  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE:  
; COMPUTER:  
; OPERATING SYSTEM:  
; SOFTWARE:  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/285,641

; FILING DATE: 1-AUGUST-1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION NUMBER: 08/076,091  
; FILING DATE: 15-JUN-1993  
; APPLICATION NUMBER: US 07/709,151  
; FILING DATE: 29-MAY-1991  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE:  
; TELEFAX:  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 20:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2213 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: genomic DNA  
; ORIGINAL SOURCE:  
; ORGANISM: Brevibacterium flavum  
; STRAIN: MJ-233  
; FEATURE:  
; NAME/KEY: promoter  
; LOCATION: 1-2213  
; IDENTIFICATION METHOD: experiment  
; US-08-285-641-20  
Alignment Scores:  
Pred. No.: 235 Length: 2213  
Score: 42.00 Matches: 9  
Percent Similarity: 70.59% Conservative: 3  
Best Local Similarity: 52.94% Mismatches: 5  
Query Match: 32.56% Indels: 0  
DB: 1 Gaps: 0  
US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-08-285-641-20 (1-2213)  
QY 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGlu 17  
|||||::: ||| ||| :::  
Db 1388 TCGGATGCTCTCAGCGCCGTGAGCAAGCAGATGCTGCTGAAATTCGA 1438  
RESULT 14  
US-08-285-641-16/c  
; Sequence 16, Application US/08285641  
; Patent No. 5726299  
; GENERAL INFORMATION:  
; APPLICANT: Zupancic, Thomas J.  
; APPLICANT: Yukawa, Hideaki  
; TITLE OF INVENTION: PROMOTER DNA FRAGMENT FROM CORNYNEFORM BACTERIA  
; NUMBER OF SEQUENCES: 24  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sughrue, Mion, Zinn, Macpeak & Seas  
; STREET: 2100 Pennsylvania Avenue, N.W.  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: United States  
; ZIP: 20037-3202  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE:  
; COMPUTER:  
; OPERATING SYSTEM:  
; SOFTWARE:  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/285,641  
; FILING DATE: 1-AUGUST-1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION NUMBER:  
; APPLICATION NUMBER: 08/076,091  
; FILING DATE: 15-JUN-1993  
; APPLICATION NUMBER: US 07/709,151  
; FILING DATE: 29-MAY-1991  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE:

Best Local Similarity:	33.33%	Mismatches:	13
Query Match:	33.33%	Indels:	0
DB:	4	Gaps:	0
US-09-847-539A-6_COPY_59_86 (1-28) x US-09-238-557-1 (1-1254)			
Qy	2	AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaVal	21
Db	499	GACATCAAGAAACCTTTGGGTGATAGTCTAGAACAATTCAGAGCAGCAAAAGCTGTATT	558
Qy	22	VallysAlaAspAsnAlaAla	28
Db	559	TTCAAGGCTGCTAAGAAGGCA	579
RESULT 10			
US-08-612-521-1/c			
; Sequence 1, Application US/08612521			
; Patent No. 5786463			
; GENERAL INFORMATION:			
; APPLICANT: Peery, Robert B			
; APPLICANT: Skatrud, Paul L			
; APPLICANT: Thornevelli, Susan J			
; TITLE OF INVENTION: MULTIPLE DRUG RESISTANCE GENE OF			
; TITLE OF INVENTION: CRYPTOCOCCUS NEOFORMANS			
; NUMBER OF SEQUENCES: 3			
; CORRESPONDENCE ADDRESS:			
; ADDRESSEE: Patent Division/AEH			
; STREET: Lilly Corporate Center			
; CITY: Indianapolis			
; STATE: Indiana			
; COUNTRY: USA			
; ZIP: 46285			
; COMPUTER READABLE FORM:			
; MEDIUM TYPE: Floppy disk			
; COMPUTER: IBM PC compatible			
; OPERATING SYSTEM: PC-DOS/MS-DOS			
; SOFTWARE: PatentIn Release #1.0, Version #1.30			
; CURRENT APPLICATION DATA:			
; APPLICATION NUMBER: US/08/612,521			
; FILING DATE:			
; CLASSIFICATION: 435			
; ATTORNEY/AGENT INFORMATION:			
; NAME: Hamilton, Amy E			
; REGISTRATION NUMBER: 33,894			
; REFERENCE/DOCKET NUMBER: X-9693			
; TELECOMMUNICATION INFORMATION:			
; TELEPHONE: 317-276-3169			
; TELEFAX: 317-276-1294			
; INFORMATION FOR SEQ ID NO: 1:			
; LENGTH: 4224 base pairs			
; TYPE: nucleic acid			
; STRANDEDNESS: double			
; TOPOLOGY: linear			
; FEATURE:			
; NAME/KEY: CDS			
; LOCATION: 1..4224			
US-08-612-521-1			
Alignment Scores:			
Pred. No.: 347 Length: 4224			
Score: 43.00 Matches: 9			
Percent Similarity: 66.67% Conservative: 3			
Best Local Similarity: 50.00% Mismatches: 6			
Query Match: 33.33% Indels: 0			
DB: 1 Gaps: 0			
US-09-847-539A-6_COPY_59_86 (1-28) x US-08-612-521-1 (1-4224)			
Qy	9	AspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVallysAlaAspAsn	26
Db	2970	GATACAAAGCATACCGATCATCAGACGAGGACCGTACGCAAGCCGATAAT	2917

```
; APPLICATION NUMBER: 08/934,846
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Dickinson, Todd Q
; REGISTRATION NUMBER: 28,354
; REFERENCE/DOCKET NUMBER: GM10088
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-994-2252
; TELEFAX: 215-994-2222
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1221 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-09-238-557-3
;
Alignment Scores:
Pred. No.: 74.5 Length: 1221
Score: 43.00 Matches: 9
Percent Similarity: 51.85% Conservatives: 5
Best Local Similarity: 33.33% Mismatches: 13
Query Match: 33.33% Indels: 0
DB: 4 Gaps: 0
;
US-09-847-539A-6_COPY_59_86 (1-28) x US-09-238-557-3 (1-1221)
Qy 2 AspalalaLeuGluAlaLeuAlaAspGlnThrAspalalaLeuGlnSerGluGluAlaAlaVal 21
Db 499 GACCATCAAGAAACCTTGGGTGATAGTCTAGAGCAATTGCAGAGCAGAGAAAGCTGGTATT 558
Qy 22 VallysAlaAspAsnAlaAla 28
Db 559 TTCAAGGCTGTTAAGAGGCA 579
;
RESULT 8
US-08-934-846-1
; Sequence 1, Application US/08934846
; Patent No. 5882898
; GENERAL INFORMATION:
; APPLICANT: Pearson, Stewart C.
; APPLICANT: Greenwood, Rebecca C.
; TITLE OF INVENTION: NOVEL folic
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
; COUNTRY: US
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; FILING DATE:
; APPLICATION NUMBER: US/08/934,846
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/934,846
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Dickinson, Todd Q
; REGISTRATION NUMBER: 28,354
; REFERENCE/DOCKET NUMBER: GM10088
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-994-2252
; TELEFAX: 215-994-2222
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1254 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-09-238-557-1
;
Alignment Scores:
Pred. No.: 77 Length: 1254
Score: 43.00 Matches: 9
Percent Similarity: 51.85% Conservatives: 5
;
US-09-847-539A-6_COPY_59_86 (1-28) x US-08-934-846-1 (1-1254)
Qy 2 AspalalaLeuGluAlaLeuAlaAspGlnThrAspalalaLeuGlnSerGluGluAlaAlaVal 21
Db 499 GACCATCAAGAAACCTTGGGTGATAGTCTAGAGCAATTGCAGAGCAGAGAAAGCTGGTATT 558
Qy 22 VallysAlaAspAsnAlaAla 28
Db 559 TTCAAGGCTGTTAAGAGGCA 579
;
RESULT 9
US-09-238-557-1
; Sequence 1, Application US/09238557
; Patent No. 6163472
; GENERAL INFORMATION:
; APPLICANT: Pearson, Stewart C.
; APPLICANT: Greenwood, Rebecca C.
; TITLE OF INVENTION: NOVEL folic
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
; COUNTRY: US
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; FILING DATE:
; APPLICATION NUMBER: US/09/238,557
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/934,846
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Dickinson, Todd Q
; REGISTRATION NUMBER: 28,354
; REFERENCE/DOCKET NUMBER: GM10088
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-994-2252
; TELEFAX: 215-994-2222
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1254 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-09-238-557-1
;
Alignment Scores:
Pred. No.: 77 Length: 1254
Score: 43.00 Matches: 9
Percent Similarity: 51.85% Conservatives: 5
```

```
; APPLICANT: Lavallie, Edward R.
; APPLICANT: Racie, Lisa A.
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Spaulding, Vikki
; APPLICANT: Agostino, Michael J.
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: MA
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/960,022
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Sprunger, Suzanne A.
; REGISTRATION NUMBER: 41,323
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8284
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1772 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-960-022-13

Alignment Scores:
Pred. No.: 78.4 Length: 1772
Score: 44.00 Matches: 9
Percent Similarity: 68.00% Conservative: 8
Best Local Similarity: 36.00% Mismatches: 8
Query Match: 34.11% Indels: 0
DB: 2 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x US-08-960-022-13 (1-1772)
QY 3 AlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaVal 22
Db 795 AGCATCAAGGCACAAAGGCGGACACAGCTGCTGAAGGAGAGGAGGCGCGCTTGTG 854
QY 23 LysAlaAspAsnAla 27
Db 855 GAAAGTACACGCA 869

RESULT 6
US-08-934-846-3
; Sequence 3, Application US/08934846
; Patent No. 5882898
; GENERAL INFORMATION:
; APPLICANT: Pearson, Stewart C.
; APPLICANT: Greenwood, Rebecca C.
; TITLE OF INVENTION: NOVEL folc
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
; COUNTRY: US
```

```
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/934,846
; FILING DATE:
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Dickinson, Todd Q
; REGISTRATION NUMBER: 28,354
; REFERENCE/DOCKET NUMBER: GM10088
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-994-2252
; TELEFAX: 215-994-2222
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1221 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-934-846-3

Alignment Scores:
Pred. No.: 74.5 Length: 1221
Score: 43.00 Matches: 9
Percent Similarity: 51.85% Conservative: 5
Best Local Similarity: 33.33% Mismatches: 13
Query Match: 33.33% Indels: 0
DB: 2 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x US-08-934-846-3 (1-1221)
QY 2 AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaVal 21
Db 499 GACCATCAAGAAACCTTGGTGATCTCTAGAAGCAATTGCAGACGACGAAAGCTGTATT 558
QY 22 ValLysAlaAspAsnAlaAla 28
Db 559 TTCAAGGCTGGTAAAGGCA 579

RESULT 7
US-09-238-557-3
; Sequence 3, Application US/09238557
; Patent No. 6165472
; GENERAL INFORMATION:
; APPLICANT: Pearson, Stewart C.
; APPLICANT: Greenwood, Rebecca C.
; TITLE OF INVENTION: NOVEL folc
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
; COUNTRY: US
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/238,557
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
```



```
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Alignment Scores:
Pred. No.: 2.94e+05 Length: 4403765
Score: 47.00 Matches: 9
Percent Similarity: 54.55% Conservative: 3
Best Local Similarity: 40.91% Mismatches: 10
Query Match: 36.43% Indels: 0
DB: 4 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x US-09-103-840A-2 (1-4403765)

Qy 7 LeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLysAlaAspAsn 26
Db 4036256 TTGATGGACGCCCGACCTGCTGGAGGCGAAGGTAAACCATCGTATCGACAAAC 4036197

Qy 27 AlaAla 28
Db 4036196 GGTGCT 4036191

RESULT 3
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Alignment Scores:
Pred. No.: 2.94e+05 Length: 4411529
Score: 47.00 Matches: 9
Percent Similarity: 54.55% Conservative: 3
Best Local Similarity: 40.91% Mismatches: 10
Query Match: 36.43% Indels: 0
DB: 4 Gaps: 0
```

```
US-09-847-539A-6_COPY_59_86 (1-28) x US-09-103-840A-1 (1-4411529)

Qy 7 LeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLysAlaAspAsn 26
Db 4044185 TTGATGGACGCCCGACCTGCTGGAGGCGAAGGTAAACCATCGTATCGACAAAC 4044126

Qy 27 AlaAla 28
Db 4044125 GGTGCT 4044120

RESULT 4
US-08-961-083-191
; Sequence 191, Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,083
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 191:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1033 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-961-083-191

Alignment Scores:
Pred. No.: 40.2 Length: 1033
Score: 44.00 Matches: 11
Percent Similarity: 65.00% Conservative: 2
Best Local Similarity: 55.00% Mismatches: 7
Query Match: 34.11% Indels: 0
DB: 3 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x US-08-961-083-191 (1-1033)

Qy 8 AlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLysAlaAspAsnAla 27
Db 8 GCAGGGCAGACAGATGCTCCCAATTGAAAGGCGCGAGTTAGCCACGAGGAGGAAGCA 67

RESULT 5
US-08-960-022-13
; Sequence 13, Application US/08960022
; Patent No. 5976837
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: MCCOY, John M.
```

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 13, 2002, 02:17:36 : Search time 9.13369 Seconds  
(without alignments)  
753.008 Million cell updates/sec

Title: US-09-847-539a-6\_COPY\_59\_86  
Perfect score: 129  
Sequence: 1 SDALEALADQTDALQSEEAAYVKADNAA 28

Scoring table: BLOSUM62  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters: -MODE=frame\_p2n.model -DEV=xlh

-Q=/cgn2\_1/USPTO\_spool/US09847539/runat\_10102002\_092549\_3456/app\_query.fasta\_1.526  
-DB=Issued\_Patents\_NA -QFWT=Fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=40  
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09847539.ecgn\_1\_13 @runat\_10102002\_092549\_3456 -NCPU=6 -ICPU=3  
-NO\_XLIFY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=60 -YGAPEXT=60 -DSELP=6 -DELEXT=7

Database : Issued\_Patents\_NA.\*  
1: /cgn2\_6/ptodata/2/lna/5A.COMB.seq.\*  
2: /cgn2\_6/ptodata/2/lna/5B.COMB.seq.\*  
3: /cgn2\_6/ptodata/2/lna/6A.COMB.seq.\*  
4: /cgn2\_6/ptodata/2/lna/6B.COMB.seq.\*  
5: /cgn2\_6/ptodata/2/lna/PCTUS.COMB.seq.\*  
6: /cgn2\_6/ptodata/2/lna/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	47	36.4	59065	4 US-09-813-817-3	Sequence 3, Appli
C 2	47	36.4	4403765	4 US-09-103-840A-2	Sequence 2, Appli
C 3	47	36.4	4411529	4 US-09-103-840A-1	Sequence 1, Appli
C 4	44	34.1	1033	3 US-08-961-083-191	Sequence 191, App
5	44	34.1	1772	2 US-08-960-022-13	Sequence 13, Appl
6	43	33.3	1221	2 US-08-934-846-3	Sequence 3, Appli
7	43	33.3	1221	4 US-09-238-557-3	Sequence 3, Appli
8	43	33.3	1254	2 US-08-934-846-1	Sequence 1, Appli
9	43	33.3	1254	4 US-09-238-557-1	Sequence 1, Appli
C 10	43	33.3	4224	1 US-08-612-521-1	Sequence 1, Appli
C 11	42	32.6	621	2 US-08-943-915-4	Sequence 4, Appli
C 12	42	32.6	1903	3 US-08-581-148C-15	Sequence 15, Appl

13	42	32.6	2213	1 US-08-285-641-20	Sequence 20, Appl
14	42	32.6	2248	1 US-08-285-641-16	Sequence 16, Appl
15	42	32.6	4447	4 US-09-521-668B-17	Sequence 17, Appl
16	42	32.6	4447	4 US-09-521-668B-19	Sequence 19, Appl
17	42	32.6	15664	1 US-08-402-282-3	Sequence 3, Appli
18	42	32.6	15664	1 US-08-508-004-3	Sequence 3, Appli
19	42	32.6	15664	1 US-08-402-066-3	Sequence 3, Appli
20	42	32.6	15664	1 US-08-402-068-3	Sequence 3, Appli
21	42	32.6	68750	3 US-09-335-409-1	Sequence 1, Appli
22	42	32.6	68750	4 US-09-568-102-1	Sequence 1, Appli
23	42	32.6	68750	4 US-09-567-969-1	Sequence 1, Appli
24	42	32.6	68750	4 US-09-568-480-1	Sequence 1, Appli
25	42	32.6	68750	4 US-09-568-486-1	Sequence 1, Appli
26	42	32.6	68750	4 US-09-568-472-1	Sequence 1, Appli
27	41	31.8	1129	5 PCT-US93-01676A-3	Sequence 3, Appli
28	41	31.8	1185	1 US-07-876-280-8	Sequence 8, Appli
29	41	31.8	1185	1 US-08-049-783-5	Sequence 5, Appli
30	41	31.8	1185	1 US-08-316-301A-9	Sequence 9, Appli
31	41	31.8	1185	2 US-08-904-278-9	Sequence 9, Appli
32	41	31.8	1185	4 US-09-076-137-9	Sequence 9, Appli
33	41	31.8	1185	4 US-09-222-594-9	Sequence 9, Appli
34	41	31.8	1185	5 PCT-US92-03624-9	Sequence 9, Appli
35	41	31.8	1242	1 US-08-252-966B-13	Sequence 13, Appl
36	41	31.8	1680	1 US-08-140-729A-2	Sequence 2, Appli
37	41	31.8	1680	1 US-08-546-666-2	Sequence 2, Appli
38	41	31.8	1680	2 US-08-916-743-2	Sequence 2, Appli
39	41	31.8	1680	2 US-09-042-929-2	Sequence 2, Appli
C 40	41	31.8	1680	2 US-08-546-661-2	Sequence 2, Appli
C 41	41	31.8	1680	2 US-09-042-960-2	Sequence 2, Appli
C 42	41	31.8	1680	3 US-09-198-650-2	Sequence 2, Appli
C 43	41	31.8	1680	3 US-09-042-913-2	Sequence 2, Appli
C 44	41	31.8	1680	3 US-09-042-937-2	Sequence 2, Appli
C 45	41	31.8	12565	4 US-09-345-217-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1  
US-09-813-817-3/c  
; Sequence 3, Application US/09813817  
; Patent No. 6340583  
; GENERAL INFORMATION:  
; APPLICANT: YAN, Chunhua et al.  
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC  
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES  
; FILE OF INVENTION: THEREOF  
; FILE REFERENCE: CLO01178  
; CURRENT APPLICATION NUMBER: US/09-813,817  
; CURRENT FILING DATE: 2001-03-22  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: PatsSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 59065  
; TYPE: DNA  
; ORGANISM: Human  
US-09-813-817-3

Alignment Scores:  
Pred. No.: 1.77e+03 Length: 59065  
Score: 47.00 Matches: 10  
Percent Similarity: 76.47% Conservative: 3  
Best Local Similarity: 58.82% Mismatches: 4  
Query Match: 36.43% Indels: 0  
DB: 4 Gaps: 0  
US-09-847-539a-6\_COPY\_59\_86 (1-28) x US-09-813-817-3 (1-59065)  
QY 8 AlaAspGlnThrAspAlaLeuGlnSerGluAlaAlaValValysala 24  
|||||  
DB 21239 GCTGAACAACTGACTCTTAGAACACGCTGAGGACAGGCTCTTCAAGCG 21189  
|||||  
RESULT 2  
US-09-103-840A-2/c

;  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Sutton, Jeffrey A.  
; REGISTRATION NUMBER: 34028  
; REFERENCE/DOCKET NUMBER: P50164  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (215) 270-5024  
; TELEFAX: (215) 270-5090  
; INFORMATION FOR SEQ ID NO: 12:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 58 amino acids  
; TYPE: amino acid  
; TOPOLOGY: circular  
; MOLECULE TYPE: peptide  
; PCT-US94-06655-12

Query Match 30.2%; Score 39; DB 5; Length 58;  
Best Local Similarity 56.2%; Pred. No. 28;  
Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 6 ALADQTDALQSEAAV 21  
|| : ||| | ||:  
Db 39 ALNQEVDALQEEVAAL 54

RESULT 39  
US-08-961-858-6  
; Sequence 6, Application US/08961858  
; Patent No. 5834210  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Shigui  
; TITLE OF INVENTION: STABLE TROPONIN SUBUNITS AND COMPLEXES  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Klauber & Jackson  
; STREET: 411 Hackensack Avenue, 4th Floor  
; CITY: Hackensack  
; STATE: New Jersey  
; COUNTRY: USA  
; ZIP: 07601  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/961,858  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Jackson Esq., David A.  
; REGISTRATION NUMBER: 26,742  
; REFERENCE/DOCKET NUMBER: 1112-1-044 CIP  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 201-487-5800  
; TELEFAX: 201-343-1684  
; TELEX: 133521  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 288 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
US-08-961-858-6

Query Match 30.2%; Score 39; DB 2; Length 288;  
Best Local Similarity 32.0%; Pred. No. 1.9e+02;  
Matches 8; Conservative 8; Mismatches 9; Indels 0; Gaps 0;

Qy 4 LEALADQTDALQSEAAVVKADNAA 28  
:| : : : |||| : ||  
Db 4 IEEVVVEEYEEQEAAVEQEAA 28

RESULT 40  
US-09-089-593-6  
; Sequence 6, Application US/09089593  
; Patent No. 6060278  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Shigui  
; TITLE OF INVENTION: STABLE TROPONIN SUBUNITS AND COMPLEXES  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Klauber & Jackson  
; STREET: 411 Hackensack Avenue, 4th Floor  
; CITY: Hackensack  
; STATE: New Jersey  
; COUNTRY: USA  
; ZIP: 07601  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/089,593  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/961,858  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Jackson Esq., David A.  
; REGISTRATION NUMBER: 26,742  
; REFERENCE/DOCKET NUMBER: 1112-1-044 CIP  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 201-487-5800  
; TELEFAX: 201-343-1684  
; TELEX: 133521  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 288 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
US-09-089-593-6

Query Match 30.2%; Score 39; DB 3; Length 288;  
Best Local Similarity 32.0%; Pred. No. 1.9e+02;  
Matches 8; Conservative 8; Mismatches 9; Indels 0; Gaps 0;

Qy 4 LEALADQTDALQSEAAVVKADNAA 28  
:| : : : |||| : ||  
Db 4 IEEVVVEEYEEQEAAVEQEAA 28

Search completed: October 13, 2002, 02:12:51  
Job time : 6.34225 secs



;; TITLE OF INVENTION: Mad or Max  
;; NUMBER OF SEQUENCES: 19  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Christensen, O'Connor, Johnson, and KindnessPLLC  
;; STREET: 1420 Fifth Ave., Suite 2800  
;; CITY: Seattle  
;; STATE: Washington  
;; COUNTRY: USA  
;; ZIP: 98101-2347  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/252.966B  
;; FILING DATE: 01-JUN-1994  
;; CLASSIFICATION: 435  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Shelton, Dennis K.  
;; REGISTRATION NUMBER: 26,997.  
;; REFERENCE/DOCKET NUMBER: FHCRI7694  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (206) 682-8100  
;; TELEFAX: (206) 224-0779  
;; INFORMATION FOR SEQ ID NO: 12:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 1253 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
;; DESCRIPTION: translation of maina cdna; see Figure 23  
;; HYPOTHEICAL: YES  
;; ORIGINAL SOURCE:  
;; ORGANISM: Mus musculus  
;; US-08-252-966B-12

Query Match 31.0%; Score 40; DB 1; Length 1253;  
Best Local Similarity 39.1%; Pred. No. 7.6e+02;  
Matches 9; Conservative 4; Mismatches 10; Indels 0; Gaps 0;

QY 4 LEALADQTDALQSEEAAYVVKADN 26  
||| : | : || : : ||  
Db 652 LEAIQKILSLSAEQAKFRLDN 674

RESULT 33  
US-07-920-281C-3  
; Sequence 3, Application US/07920281C  
; Patent No. 5739026  
; GENERAL INFORMATION:  
; APPLICANT: Garoff, Henrik  
; APPLICANT: Liljestrom, Peter  
; TITLE OF INVENTION: DNA Expression Systems Based on  
; NUMBER OF SEQUENCES: 27  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Birch, Stewart, Kolasch & Birch  
; STREET: P.O. Box 747  
; STATE: Virginia  
; COUNTRY: USA  
; ZIP: 22040-0747  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/920.281C  
; FILING DATE: 13-AUG-1992  
; CLASSIFICATION: 435

;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Murphy Jr., Gerald M.  
;; REGISTRATION NUMBER: 28,977  
;; REFERENCE/DOCKET NUMBER: 828-103P  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 703-241-1300  
;; TELEFAX: 703-241-2848  
;; TELEX: 248345  
;; INFORMATION FOR SEQ ID NO: 3:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 1253 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
;; SEQUENCE DESCRIPTION: SEQ ID NO: 3:  
US-07-920-281C-3

Query Match 31.0%; Score 40; DB 1; Length 1253;  
Best Local Similarity 33.3%; Pred. No. 7.6e+02;  
Matches 8; Conservative 6; Mismatches 10; Indels 0; Gaps 0;

QY 5 EALADQTDALQSEEAAYVVKADNAA 28  
|| :| : : | : || :  
Db 920 EAYVDRSDVCRHDSAYKAHTAS 943

RESULT 34  
US-08-466-277-3  
; Sequence 3, Application US/08466277  
; Patent No. 6190666  
; GENERAL INFORMATION:  
; APPLICANT: Garoff, Henrik  
; APPLICANT: Liljestrom, Peter  
; TITLE OF INVENTION: DNA Expression Systems Based on  
; NUMBER OF SEQUENCES: 27  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Birch, Stewart, Kolasch & Birch  
; STREET: P.O. Box 747  
; CITY: Falls Church  
; STATE: Virginia  
; COUNTRY: USA  
; ZIP: 22040-0747  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/466,277  
; FILING DATE: 06-Jun-1995  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/920,281  
; FILING DATE: <Unknown>  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Murphy Jr., Gerald M.  
; REGISTRATION NUMBER: 28,977  
; REFERENCE/DOCKET NUMBER: 828-103P  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 703-241-1300  
; TELEFAX: 703-241-2848  
; TELEX: 248345  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1253 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; SEQUENCE DESCRIPTION: SEQ ID NO: 3:  
US-08-466-277-3

Query Match 31.0%; Score 40; DB 4; Length 1253;  
Best Local Similarity 33.3%; Pred. No. 7.6e+02;

```
QY 1 SDALEALADQTDALQSEAAV 21
: | | : | | : | | |
Db 469 SDVLVFTSEQCDHRSEAAV 489

RESULT 29
US-08-630-822A-62
; Sequence 62, Application US/08630822A
; Patent No. 5840695
; GENERAL INFORMATION:
; APPLICANT: FRANK, GLENN R.
; APPLICANT: HUNTER, SHIRLEY WU
; APPLICANT: WALLENFELS, LYNDIA
; TITLE OF INVENTION: NOVEL ECTOPARASITE SALIVA PROTEINS
; NUMBER OF SEQUENCES: 107
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross P.C.
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: U.S.A.
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/630,822A
; FILING DATE: 11-APR-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: CONNELL, GARY J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2618-17-C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 62:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 900 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-630-822A-62

Query Match 31.0%; Score 40; DB 2; Length 900;
Best Local Similarity 40.0%; Pred. No. 5.1e+02;
Matches 8; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

QY 2 DALEALADQTDALQSEAAV 21
: : : | | | | | :
Db 536 ETMQLQTDALQSEAAV 555

RESULT 30
US-09-005-069-62
; Sequence 62, Application US/09005069
; Patent No. 5932470
; GENERAL INFORMATION:
; APPLICANT: FRANK, GLENN R.
; APPLICANT: HUNTER, SHIRLEY WU
; APPLICANT: WALLENFELS, LYNDIA
; TITLE OF INVENTION: NOVEL ECTOPARASITE SALIVA PROTEINS
; NUMBER OF SEQUENCES: 107
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross P.C.
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: U.S.A.
```

```
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/005,069
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/630,822
FILING DATE: 11-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: CONNELL, GARY J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2618-17-C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 62:
SEQUENCE CHARACTERISTICS:
LENGTH: 900 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-005-069-62

Query Match 31.0%; Score 40; DB 2; Length 900;
Best Local Similarity 40.0%; Pred. No. 5.1e+02;
Matches 8; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

QY 2 DALEALADQTDALQSEAAV 21
: : : | | | | | :
Db 536 ETMQLQTDALQSEAAV 555

RESULT 31
US-09-268-347-32
; Sequence 32, Application US/09268347
; Patent No. 6335182
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; TITLE OF INVENTION: RECOMBINANT HAEMOPHILUS INFLUENZAE ADHESIN PROTEINS
; FILE REFERENCE: 1038-860
; CURRENT APPLICATION NUMBER: US/09/268,347
; CURRENT FILING DATE: 1999-03-16
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 32
; LENGTH: 1094
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-268-347-32

Query Match 31.0%; Score 40; DB 4; Length 1094;
Best Local Similarity 37.5%; Pred. No. 6.5e+02;
Matches 9; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

QY 3 ALEALADQTDALQSEAAVVKADN 26
: | | | | | :
Db 811 AKAFGDETKALSSDKLETYNAND 834

RESULT 32
US-08-252-966B-12
; Sequence 12, Application US/08252966B
; Patent No. 5624818
; GENERAL INFORMATION:
; APPLICANT: Eisenman, Robert N.
; APPLICANT: Hurlin, Peter J.
; APPLICANT: Aver, Donald E.
; TITLE OF INVENTION: Regulatory Proteins that Dimerize with
```

```
QY      5  EALADQTDALQSEAAVVKADN 26
      | | : | | : | | : | | : |
Db     231 EELKKVDDLKKELEAAIRAKEN 252

RESULT 26
US-08-858-207A-394
; Sequence 394, Application US/08858207A
; Patent No. 6348328
; GENERAL INFORMATION:
; APPLICANT: Black, Michael
; APPLICANT: Hodgson, John
; APPLICANT: Knowles, David
; APPLICANT: Nicholas, Richard
; APPLICANT: Stodola, Robert
; TITLE OF INVENTION: No. 6348328e1 Compounds
; NUMBER OF SEQUENCES: 552
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/858,207A
; FILING DATE: 09-MAY-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/017670
; FILING DATE: 14-MAY-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Glimi, Edward R
; REGISTRATION NUMBER: 38,891
; REFERENCE/DOCKET NUMBER: P50475
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-4478
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 394:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 119 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 6348328e
US-08-858-207A-394

Query Match      31.0%; Score 40; DB 4; Length 119;
Best Local Similarity 30.8%; Pred. No. 47;
Matches 8; Conservative 8; Mismatches 10; Indels 0; Gaps 0;

QY      2  DALEALADQTDALQSEAAVVKADNA 27
      | | : | | : | | : | | : |
Db     8  DKVDALLDQLSANSKDKTALLESTKA 33

RESULT 27
US-08-457-245-3
; Sequence 3, Application US/08457245
; Patent No. 5573915
; GENERAL INFORMATION:
; APPLICANT: BARRY III, Clifton E.
; APPLICANT: YUAN, Ying
; TITLE OF INVENTION: CLONING AND EXPRESSION OF DNA INVOLVED
; IN THE BIOSYNTHESIS OF CYCLOPROPANATED MYCOLIC ACIDS IN
; MYCOBACTERIUM TUBERCULOSIS
; TITLE OF INVENTION:

QY      1  SDALEALADQTDALQSEEE 18
      | | | | | | | | | | | |
Db     240 SAALQANKQAIALQSEEE 257

RESULT 28
US-09-245-041-9
; Sequence 9, Application US/09245041
; Patent No. 6274339
; GENERAL INFORMATION:
; APPLICANT: Moore, K.
; APPLICANT: Nagle, D.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE DIAGNOSIS AND TREATMENT
; OF BODY WEIGHT DISORDERS INCLUDING OBESITY
; FILE REFERENCE: 7853-136
; CURRENT APPLICATION NUMBER: US/09/245,041
; CURRENT FILING DATE: 1999-02-05
; EARLIER APPLICATION NUMBER: 60/093,630
; EARLIER FILING DATE: 1998-07-21
; EARLIER APPLICATION NUMBER: 60/104,978
; EARLIER FILING DATE: 1998-10-20
; NUMBER OF SEQ ID NOS: 131
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 9
; LENGTH: 549
; TYPE: PRT
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: SITE
; LOCATION: all Xaa positions
; OTHER INFORMATION: Xaa-unknown amino acid
US-09-245-041-9

Query Match      31.0%; Score 40; DB 4; Length 549;
Best Local Similarity 47.6%; Pred. No. 2.9e+02;
Matches 10; Conservative 3; Mismatches 8; Indels 0; Gaps 0;
```

;; TITLE OF INVENTION: NO. 6166195el Nematode-Active Toxins and Genes Which Code  
;; FILE OF INVENTION: Therefor  
;; CURRENT APPLICATION NUMBER: MA-20CCCD2  
;; CURRENT FILING DATE: 1998-05-12  
;; EARLIER APPLICATION NUMBER: 08/316,301  
;; EARLIER FILING DATE: 1994-09-30  
;; NUMBER OF SEQ ID NOS: 42  
;; SOFTWARE: PatentIn Ver. 2.1  
;; SEQ ID NO 10  
;; LENGTH: 395  
;; TYPE: PRT  
;; ORGANISM: Bacillus thuringiensis  
US-09-076-137-10

Query Match 31.8%; Score 41; DB 4; Length 395;  
Best Local Similarity 40.9%; Pred. No. 1.4e+02;  
Matches 9; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

QY 5 EALADQTDALQSEEAHVKADN 26  
DB 231 EELLKKVDDLKKELEAAIKAE 252

## RESULT 24

;; Sequence 10, Application US/09222594  
;; Patent No. 6303364

## ;; GENERAL INFORMATION:

;; APPLICANT: Thompson, Mark  
;; APPLICANT: Knuth, Mark  
;; APPLICANT: Cardineau, Guy

;; TITLE OF INVENTION: Bacillus thuringiensis Toxins with Improved

;; TITLE OF INVENTION: Activity

;; NUMBER OF SEQUENCES: 10

;; CORRESPONDENCE ADDRESS:

;; ADDRESSEE: Saliwanchik & Saliwanchik

;; STREET: 2421 N.W. 41st Street, Suite A-1

;; CITY: Gainesville

;; STATE: Florida

;; COUNTRY: USA

;; ZIP: 32606

;; COMPUTER READABLE FORM:

;; MEDIUM TYPE: Floppy disk

;; COMPUTER: IBM PC compatible

;; OPERATING SYSTEM: PC-DOS/MS-DOS

;; SOFTWARE: PatentIn

;; CURRENT APPLICATION DATA:

;; APPLICATION NUMBER: US/09/222,594

;; FILING DATE:

;; CLASSIFICATION:

;; PRIOR APPLICATION DATA:

;; APPLICATION NUMBER: 08/904,278

;; FILING DATE:

;; ATTORNEY/AGENT INFORMATION:

;; NAME: Sanders, Jay M.

;; REGISTRATION NUMBER: 39,355

;; REFERENCE/DOCKET NUMBER: MA-702

;; TELECOMMUNICATION INFORMATION:

;; TELEPHONE: (352) 375-8100

;; TELEFAX: (352) 372-5800

;; INFORMATION FOR SEQ ID NO: 10:

;; SEQUENCE CHARACTERISTICS:

;; LENGTH: 395 amino acids

;; TYPE: amino acid

;; STRANDEDNESS: single

;; TOPOLOGY: linear

;; MOLECULE TYPE: protein

;; HYPOTHETICAL: YES

;; ANTI-SENSE: NO

;; ORIGINAL SOURCE:

;; ORGANISM: BACILLUS THURINGIENSIS

;; INDIVIDUAL ISOLATE: PS69D1

;; IMMEDIATE SOURCE:  
;; CLONE: E. coli NM522(pMYC2317) NRRL B-18816  
;; FEATURE:  
;; NAME/KEY: Protein  
;; LOCATION: 1..395  
US-09-222-594-10

Query Match 31.8%; Score 41; DB 4; Length 395;  
Best Local Similarity 40.9%; Pred. No. 1.4e+02;  
Matches 9; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

QY 5 EALADQTDALQSEEAHVKADN 26  
DB 231 EELLKKVDDLKKELEAAIKAE 252

## RESULT 25

;; Sequence 10, Application PC/TUS9203624

;; GENERAL INFORMATION:

;; APPLICANT: Schnepf, Harry E.

;; APPLICANT: Schwab, George E.

;; APPLICANT: Payne, Jewel M.

;; APPLICANT: Narva, Kenneth E.

;; APPLICANT: Poncarrada, Luis

;; TITLE OF INVENTION: Novel Nematode-Active Toxins and Genes

;; TITLE OF INVENTION: Which Code Therefor

;; NUMBER OF SEQUENCES: 40

;; CORRESPONDENCE ADDRESS:

;; ADDRESSEE: David R. Saliwanchik

;; STREET: 2421 N.W. 41st Street, Suite A-1

;; CITY: Gainesville

;; STATE: FL

;; COUNTRY: USA

;; ZIP: 32606

;; COMPUTER READABLE FORM:

;; MEDIUM TYPE: Floppy disk

;; COMPUTER: IBM PC compatible

;; OPERATING SYSTEM: PC-DOS/MS-DOS

;; SOFTWARE: PatentIn Release #1.0, Version #1.25

;; CURRENT APPLICATION DATA:

;; APPLICATION NUMBER: PCT/US92/03624

;; FILING DATE: 19920501

;; CLASSIFICATION: 435

;; ATTORNEY/AGENT INFORMATION:

;; NAME: Saliwanchik, David R.

;; REGISTRATION NUMBER: 31,794

;; REFERENCE/DOCKET NUMBER: MA20C2C1C1

;; TELECOMMUNICATION INFORMATION:

;; TELEPHONE: 904-375-8100

;; TELEFAX: 904-372-5800

;; INFORMATION FOR SEQ ID NO: 10 (PS69D1):

;; SEQUENCE CHARACTERISTICS:

;; LENGTH: 395 amino acids

;; TYPE: AMINO ACID

;; STRANDEDNESS: single

;; TOPOLOGY: linear

;; MOLECULE TYPE: protein

;; HYPOTHETICAL: YES

;; ANTI-SENSE: NO

;; ORIGINAL SOURCE:

;; ORGANISM: BACILLUS THURINGIENSIS

;; INDIVIDUAL ISOLATE: PS69D1

;; IMMEDIATE SOURCE:

;; CLONE: E. coli NM522(pMYC2317) NRRL B-18816

;; FEATURE:

;; NAME/KEY: Protein

;; LOCATION: 1..395

;; PCT-US92-03624-10

Query Match 31.8%; Score 41; DB 5; Length 395;  
Best Local Similarity 40.9%; Pred. No. 1.4e+02;  
Matches 9; Conservative 4; Mismatches 9; Indels 0; Gaps 0;





TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US93-01676A-4

Query Match 31.8%; Score 41; DB 5; Length 370;  
Best Local Similarity 43.5%; Pred. No. 1.3e+02;  
Matches 10; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

QY 1 SDALALADQTDALQSEEAAYVK 23  
Db 343 ADSLMLARQVSRLESQGAALPK 365

## RESULT 19

US-07-876-280-9  
Sequence 9, Application US/07876280  
Patent No. 5262158  
GENERAL INFORMATION:  
APPLICANT: Payne, Jewel M.  
APPLICANT: Cannon, Raymond J.C.  
APPLICANT: Bagley, Angela L.  
TITLE OF INVENTION: No. 5262158el Bacillus thuringiensis Isolates for  
CONTROLLING ACARIDES  
NUMBER OF SEQUENCES: 30  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: David R. Sallwanchik  
STREET: 2421 N.W. 41st Street, Suite A-1  
CITY: Gainesville  
STATE: FL  
COUNTRY: USA  
ZIP: 32606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
FILING DATE: 19920430  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Sallwanchik, David R.  
REGISTRATION NUMBER: 31,794  
REFERENCE/DOCKET NUMBER: M/S 104  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 904-375-8100  
TELEFAX: 904-372-5800  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 395 amino acids  
TYPE: AMINO ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: BACILLUS THURINGIENSIS  
INDIVIDUAL ISOLATE: PS69D1  
IMMEDIATE SOURCE:  
CLONE: E. coli NM522(pMYC2317) NRRL B-18816  
FEATURE:  
NAME/KEY: Protein  
LOCATION: 1..395

US-07-876-280-9  
Query Match 31.8%; Score 41; DB 1; Length 395;  
Best Local Similarity 40.9%; Pred. No. 1.4e+02;  
Matches 9; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

QY 5 EALADQTDALQSEEAAYVKADN 26

Db 231 EELKKVDDLKKELEAAIKAE 252

## RESULT 20

US-08-049-783-6  
Sequence 6, Application US/08049783  
Patent No. 5439881  
GENERAL INFORMATION:  
APPLICANT: Narva, Kenneth E  
APPLICANT: Schwab, George E  
APPLICANT: Payne, Jewel M  
TITLE OF INVENTION: Gene Encoding No. 5439881el Nematode-Active  
Toxins Cloned from Bacillus thuringiensis Isolates  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Jeff Lloyd  
STREET: 2421 N.W. 41st Street  
CITY: Gainesville  
STATE: FL  
COUNTRY: USA  
ZIP: 32606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/049,783  
FILING DATE: 19930419  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Lloyd, Jeff  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 904-375-8100  
TELEFAX: 904-372-5800  
INFORMATION FOR SEQ ID NO: 6 (PS69D1):  
SEQUENCE CHARACTERISTICS:  
LENGTH: 395 amino acids  
TYPE: AMINO ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: BACILLUS THURINGIENSIS  
INDIVIDUAL ISOLATE: PS69D1  
IMMEDIATE SOURCE:  
LIBRARY: LAMBDAGEM(TM)-11 LIBRARY OF KENNETH NARVA  
CLONE: PS69D1A  
FEATURE:  
NAME/KEY: Protein  
LOCATION: 1..395

US-08-049-783-6  
Query Match 31.8%; Score 41; DB 1; Length 395;  
Best Local Similarity 40.9%; Pred. No. 1.4e+02;  
Matches 9; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

QY 5 EALADQTDALQSEEAAYVKADN 26

Db 231 EELKKVDDLKKELEAAIKAE 252

## RESULT 21

US-08-316-301A-10  
Sequence 10, Application US/08316301A  
Patent No. 5753492  
GENERAL INFORMATION:  
APPLICANT: Schnepf, Harry E.  
APPLICANT: Schwab, George E.  
APPLICANT: Payne, Jewel M.  
APPLICANT: Narva, Kenneth E.  
APPLICANT: Foncerrada, Luis





```
;
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/764,100
; FILING DATE: 06-DEC-1996
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/214,064
; FILING DATE:
; APPLICATION NUMBER: US 08/032,235
; FILING DATE: 17-MAR-1993
; APPLICATION NUMBER: GB 9206016.9
; FILING DATE: 19-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: NO. 5773700ris, Allen E.
; REGISTRATION NUMBER: 34,490
; REFERENCE/DOCKET NUMBER: 137-1061
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 354-3592
; TELEFAX: (415) 857-1125
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 744 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-764-100-25

Query Match 33.3%; Score 43; DB 1; Length 744;
Best Local Similarity 38.5%; Pred. No. 1.5e+02;
Matches 10; Conservative 6; Mismatches 10; Indels 0; Gaps 0;

QY 1 SDALEALADQTDALQSEAAVVRADN 26
   ||| : ||||| : |||
Db 432 SERLEPRLEILEALASERVALLEASN 457

RESULT 8
US-08-710-749-8
; Sequence 8, Application US/08710749
; Patent No. 5955089
; GENERAL INFORMATION:
; APPLICANT: Brilles, David E.
; APPLICANT: Hollingshead, Susan
; APPLICANT: Becker, Robert
; TITLE OF INVENTION: STRAIN SELECTION OF PNEUMOCOCCAL SURFACE
; TITLE OF INVENTION: PROTEINS
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Curtis, Morris & Safford
; STREET: 530 Fifth Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/710,749
; FILING DATE: 20-SEP-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Frommer, William S.
; REGISTRATION NUMBER: 25,506
; REFERENCE/DOCKET NUMBER: 454312-2074
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 840-3333
; TELEFAX: (212) 840-0712

;
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 102 amino acids
; TYPE: amino acid
; STRANDEDNESS: n/a
; TOPOLOGY: linear
; MOLECULE TYPE: amino acid
; US-08-710-749-8

Query Match 32.6%; Score 42; DB 2; Length 102;
Best Local Similarity 45.5%; Pred. No. 20;
Matches 10; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY 4 LEALADQTDALQSEAAVVRADN 25
   ||| : ||| : ||| : |||
Db 36 LEELSDKIDELDAEIAKNLAKD 57

RESULT 9
US-09-521-668B-18
; Sequence 18, Application US/09521668B
; Patent No. 6303383
; GENERAL INFORMATION:
; APPLICANT: NAKAMURA, JUN
; APPLICANT: KANNO, SOHEI
; APPLICANT: KIMURA, EIICHIRO
; APPLICANT: MATSUI, KAHIKO
; APPLICANT: NAKAMATSU, TSUYOSHI
; TITLE OF INVENTION: TEMPERATURE SENSITIVE PLASMID FOR CORYNEFORM BACTERIA
; FILE REFERENCE: 0010-1093-0
; CURRENT APPLICATION NUMBER: US/09/521,668B
; CURRENT FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: JP 11-69896
; PRIOR FILING DATE: 1999-03-16
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 18
; LENGTH: 427
; TYPE: PRT
; ORGANISM: Brevibacterium lactofermentum
; US-09-521-668B-18

Query Match 32.6%; Score 42; DB 4; Length 427;
Best Local Similarity 47.4%; Pred. No. 1.1e+02;
Matches 9; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 4 LEALADQTDALQSEAAVVRADN 22
   ||| : ||| : ||| : |||
Db 256 LDMLADQSDAGEDMDAVLV 274

RESULT 10
US-09-521-668B-20
; Sequence 20, Application US/09521668B
; Patent No. 6303383
; GENERAL INFORMATION:
; APPLICANT: NAKAMURA, JUN
; APPLICANT: KANNO, SOHEI
; APPLICANT: KIMURA, EIICHIRO
; APPLICANT: MATSUI, KAHIKO
; APPLICANT: NAKAMATSU, TSUYOSHI
; TITLE OF INVENTION: TEMPERATURE SENSITIVE PLASMID FOR CORYNEFORM BACTERIA
; FILE REFERENCE: 0010-1093-0
; CURRENT APPLICATION NUMBER: US/09/521,668B
; CURRENT FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: JP 11-69896
; PRIOR FILING DATE: 1999-03-16
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 20
; LENGTH: 427
; TYPE: PRT
; ORGANISM: Brevibacterium lactofermentum
```

```
;
;
; TELEX:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 407 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-09-238-557-4
;
Query Match 33.3%; Score 43; DB 4; Length 407;
Best Local Similarity 33.3%; Pred. No. 74;
Matches 9; Conservative 5; Mismatches 13; Indels 0; Gaps 0;

Qy 2 DALEALADQTDALQSEEAHVKNAA 28
Db 167 DHQETLGDLSLEIAEQKAGIFKAGKA 193

RESULT 5
US-08-934-846-2
; Sequence 2, Application US/08934846
; Patent No. 5882898
; GENERAL INFORMATION:
; APPLICANT: Pearson, Stewart C.
; APPLICANT: Greenwood, Rebecca C.
; TITLE OF INVENTION: NOVEL FOIC
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
; COUNTRY: US
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/934,846
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/934,846
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Dickinson, Todd Q
; REGISTRATION NUMBER: 28,354
; REFERENCE/DOCKET NUMBER: GM10088
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-994-2252
; TELEFAX: 215-994-2222
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 418 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-09-238-557-2
;
Query Match 33.3%; Score 43; DB 2; Length 418;
Best Local Similarity 33.3%; Pred. No. 74;
Matches 9; Conservative 5; Mismatches 13; Indels 0; Gaps 0;

Qy 2 DALEALADQTDALQSEEAHVKNAA 28
Db 167 DHQETLGDLSLEIAEQKAGIFKAGKA 193

RESULT 6
US-09-238-557-2
;
;
; TELEX:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 407 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-09-238-557-4
;
Query Match 33.3%; Score 43; DB 4; Length 407;
Best Local Similarity 33.3%; Pred. No. 72;
Matches 9; Conservative 5; Mismatches 13; Indels 0; Gaps 0;

Qy 2 DALEALADQTDALQSEEAHVKNAA 28
Db 167 DHQETLGDLSLEIAEQKAGIFKAGKA 193
```

```
;
;
; Sequence 2, Application US/09238557
; Patent No. 6165472
; GENERAL INFORMATION:
; APPLICANT: Pearson, Stewart C.
; APPLICANT: Greenwood, Rebecca C.
; TITLE OF INVENTION: NOVEL FOIC
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
; COUNTRY: US
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/238,557
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/934,846
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Dickinson, Todd Q
; REGISTRATION NUMBER: 28,354
; REFERENCE/DOCKET NUMBER: GM10088
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-994-2252
; TELEFAX: 215-994-2222
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 418 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-09-238-557-2
;
Query Match 33.3%; Score 43; DB 4; Length 418;
Best Local Similarity 33.3%; Pred. No. 74;
Matches 9; Conservative 5; Mismatches 13; Indels 0; Gaps 0;

Qy 2 DALEALADQTDALQSEEAHVKNAA 28
Db 167 DHQETLGDLSLEIAEQKAGIFKAGKA 193

RESULT 7
US-08-764-100-25
; Sequence 25, Application US/08764100
; Patent No. 5773700
; GENERAL INFORMATION:
; APPLICANT: van Grinsven J., Martinus Q.
; APPLICANT: De Haan, Petrus T.
; APPLICANT: Gielen L., Johannes J.
; APPLICANT: Peters, Dirk
; APPLICANT: Goldbach, Robert W.
; TITLE OF INVENTION: Improvements in or Relating to Organic
; TITLE OF INVENTION: Compounds
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sandoz Agro, Inc
; STREET: 975 California Avenue
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```

```
RESULT 2
US-08-960-022-14
; Sequence 14, Application US/08960022
; Patent No. 5976837
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M.
; APPLICANT: LaVallie, Edward R.
; APPLICANT: Racie, Lisa A.
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Spaulding, Vikki
; APPLICANT: Agostino, Michael J.
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: MA
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/960,022
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Sprunger, Suzanne A.
; REGISTRATION NUMBER: 41,323
; TELEPHONE: (617) 498-8284
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 514 amino acids
; TYPE: amino acid
; STRANDEDNESS: linear
; TOPOLOGY: protein
; MOLECULE TYPE: protein
US-08-960-022-14

Query Match 34.1%; Score 44; DB 2; Length 514;
Best Local Similarity 36.0%; Pred. No. 67;
Matches 9; Conservative 8; Mismatches 8; Indels 0; Gaps 0;

QY 3 ALEALADQTDALQSEEAHVKAADNA 27
Db 240 SIRAQRQATQLLKREEARLVESNNA 264

RESULT 3
US-08-934-846-4
; Sequence 4, Application US/08934846
; Patent No. 5882898
; GENERAL INFORMATION:
; APPLICANT: Pearson, Stewart C.
; APPLICANT: Greenwood, Rebecca C.
; TITLE OF INVENTION: NOVEL folc
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
; COUNTRY: US
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/238,557
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION NUMBER: 08/934,846
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Dickinson, Todd Q
; REGISTRATION NUMBER: 28,354
; REFERENCE/DOCKET NUMBER: GM10088
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-994-2252
; TELEFAX: 215-994-2222
```

```
QY 2 DALEALADQTDALQSEEAHVKAADNA 28
Db 167 DHOETLGDSLEAIEKAGIFKAGKKA 193

RESULT 4
US-09-238-557-4
; Sequence 4, Application US/09238557
; Patent No. 6165472
; GENERAL INFORMATION:
; APPLICANT: Pearson, Stewart C.
; APPLICANT: Greenwood, Rebecca C.
; TITLE OF INVENTION: NOVEL folc
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
; COUNTRY: US
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/238,557
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION NUMBER: 08/934,846
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Dickinson, Todd Q
; REGISTRATION NUMBER: 28,354
; REFERENCE/DOCKET NUMBER: GM10088
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-994-2252
; TELEFAX: 215-994-2222

Query Match 33.3%; Score 43; DB 2; Length 407;
Best Local Similarity 33.3%; Pred. No. 72;
Matches 9; Conservative 5; Mismatches 13; Indels 0; Gaps 0;

QY 2 DALEALADQTDALQSEEAHVKAADNA 28
Db 167 DHOETLGDSLEAIEKAGIFKAGKKA 193
```

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2002, 02:06:51 ; Search time 4.34225 Seconds  
(without alignments)  
157.503 Million cell updates/sec

Title: US-09-847-539A-6\_COPY\_59\_86  
Perfect score: 129  
Sequence: 1 SDALEALADQTDALQSEAAVVKADNA 28

Scoring table: BLOSUM62  
Gapop 60.0 , Gapext 60.0

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.\*

- 1: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep.\*
- 2: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep.\*
- 3: /cgn2\_6/ptodata/2/1aa/6A\_COMB.pep.\*
- 4: /cgn2\_6/ptodata/2/1aa/6B\_COMB.pep.\*
- 5: /cgn2\_6/ptodata/2/1aa/PCTUS\_COMB.pep.\*
- 6: /cgn2\_6/ptodata/2/1aa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	44	34.1	344	4	US-08-961-083-192
2	44	34.1	514	2	US-08-960-022-14
3	43	33.3	407	2	US-08-934-846-4
4	43	33.3	407	4	US-09-238-557-4
5	43	33.3	418	2	US-08-934-846-2
6	43	33.3	418	4	US-09-238-557-2
7	43	33.3	744	1	US-08-764-100-25
8	42	32.6	102	2	US-08-710-749-8
9	42	32.6	427	4	US-09-521-668B-18
10	42	32.6	427	4	US-09-521-668B-20
11	42	32.6	2890	4	US-09-413-814-67
12	42	32.6	3798	3	US-09-335-409-6
13	42	32.6	3798	4	US-09-568-102-6
14	42	32.6	3798	4	US-09-567-969-6
15	42	32.6	3798	4	US-09-568-480-6
16	42	32.6	3798	4	US-09-568-486-6
17	42	32.6	3798	4	US-09-568-472-6
18	41	31.8	370	5	PCT-US93-01676A-4
19	41	31.8	395	1	US-07-876-280-9
20	41	31.8	395	1	US-08-049-783-6
21	41	31.8	395	1	US-08-316-301A-10
22	41	31.8	395	2	US-08-904-278-10
23	41	31.8	395	4	US-09-076-137-10
24	41	31.8	395	4	US-09-222-594-10
25	41	31.8	395	5	PCT-US92-03624-10
26	40	31.0	119	4	US-08-858-207A-394
27	40	31.0	287	1	US-08-457-245-3

28 40 31.0 549 4 US-09-245-041-9 Sequence 9, Appl  
29 40 31.0 900 2 US-08-630-822A-62 Sequence 62, Appl  
30 40 31.0 900 2 US-09-005-069-62 Sequence 62, Appl  
31 40 31.0 1094 4 US-09-268-347-32 Sequence 32, Appl  
32 40 31.0 1253 1 US-08-252-966B-12 Sequence 12, Appl  
33 40 31.0 1253 1 US-07-920-281C-3 Sequence 3, Appl  
34 40 31.0 1253 4 US-08-466-277-3 Sequence 3, Appl  
35 40 31.0 1260 4 US-09-245-041-2 Sequence 2, Appl  
36 40 31.0 1261 1 US-08-252-966B-18 Sequence 18, Appl  
37 39 30.2 25 6 5169933-42 Patent No. 5169933  
38 39 30.2 58 5 PCT-US94-06655-12 Sequence 12, Appl  
39 39 30.2 288 3 US-08-961-858-6 Sequence 6, Appl  
40 39 30.2 288 3 US-09-089-593-6 Sequence 6, Appl  
41 39 30.2 288 3 US-08-950-925-4 Sequence 4, Appl  
42 39 30.2 326 1 US-07-603-133B-25 Sequence 25, Appl  
43 39 30.2 528 4 US-08-928-213B-8 Sequence 8, Appl  
44 39 30.2 770 1 US-08-445-135-2 Sequence 2, Appl  
45 39 30.2 912 5 PCT-US95-03747-2 Sequence 2, Appl

## ALIGNMENTS

## RESULT 1

US-08-961-083-192  
; Sequence 192, Application US/08961083  
; Patent No. 6159469  
; GENERAL INFORMATION:  
; APPLICANT: Choi et. al.  
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines  
; NUMBER OF SEQUENCES: 452  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
; COMPUTER: HP Vectra 486/33  
; OPERATING SYSTEM: MSDOS version 6.2  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/961.083  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Brookes, A. Anders  
; REGISTRATION NUMBER: 36,373  
; REFERENCE/DOCKET NUMBER: PB340P2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (301) 309-8504  
; TELEFAX: (301) 309-8512  
; INFORMATION FOR SEQ ID NO: 192:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 344 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-961-083-192

Query Match 34.1%; Score 44; DB 4; Length 344;

Best Local Similarity 55.0%; Pred. No. 42;

Matches 11; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

Qy 8 ADOTDALQSEAAVVKADNA 27

Db 3 ACQTDSQIEKANVSQGGKA 22



**THIS PAGE BLANK (USPTO)**



```
QY 26 EKLALRNEERAIDELKKQAIEDKEATTAIEAASSDALEALADQTDALQSEEAHVVDNA 85
   I : : I : : : : I : : I : : I : : I : : I : : I : : I : : I : :
Db 227 EQARIREQARLAQARQAAQAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAE 286
   I : : I : : I : : I : : I : : I : : I : : I : : I : : I : : I : :
QY 86 ASDALEALADQTDALQSEEAHVVDNAASDAWEKAATPIALDVKKTKDTFVVKKEERQ 145
   I : : I : : I : : I : : I : : I : : I : : I : : I : : I : : I : :
Db 287 ESKAPESTTTTEESTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTTTEE 346
   I : : I : : I : : I : : I : : I : : I : : I : : I : : I : : I : :
QY 146 NVNTLPTTGEESN 158
   I : : I : : I : : I : : I : : I : : I : : I : : I : : I : : I : :
Db 347 STSTESTTTEES 359

RESULT 34
Q9KWR3 PRELIMINARY; PRT; 2178 AA.
ID Q9KWR3
AC Q9KWR3;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE STREPTOCOCCAL HEMAGGLUTININ.
GN HSA.
OS Streptococcus gordonii.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
ON NCBI_TaxID=1302;
RX SEQUENCE FROM N.A.
RP STRAIN-DL1;
RC Takahashi Y., Konishi K., Yoshikawa M.;
RA "Cloning and characterization of the gene encoding a hemagglutinin of
RT Streptococcus gordonii DL1.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB029393; BAA97453.1;
DR InterPro; IPR004089; Chemotaxis_transducer.
DR InterPro; IPR001899; Gram_pos_anchor.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN.1.
SQ SEQUENCE 2178 AA; 203507 MW; 75CC2E27F41DA8C CRC64;

Query Match 10.0%; Score 78; DB 2; Length 2178;
Best Local Similarity 21.4%; Pred. No. 22;
Matches 24; Conservative 27; Mismatches 61; Indels 0; Gaps 0;

QY 47 DKEATTAIEAASSDALEALADQTDALQSEEAHVVDNAASDALEALADQTDALQSEEAHV 106
   I : : I : : I : : I : : I : : I : : I : : I : : I : : I : : I : :
Db 1471 ESASTSASVSASTSASTSASVSASTSASTSASVSASTSASTSASVSASTSASVSAS 1530
   I : : I : : I : : I : : I : : I : : I : : I : : I : : I : : I : :
QY 107 VQSDNAASDAWEKAATPIALDVKKTKDTFVVKKEERQNTLPTTGEESN 158
   I : : I : : I : : I : : I : : I : : I : : I : : I : : I : : I : :
Db 1531 TSASTSASVSASEASTSASVSASEASTSASVSASEASTSASVSASTSASVSASTSAS 1582

RESULT 35
Q9U4R2 PRELIMINARY; PRT; 310 AA.
ID Q9U4R2
AC Q9U4R2;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE ANON1A3 (FRAGMENT).
GN CG1910 OR ANON1A3.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
ON NCBI_TaxID=7227;
RX SEQUENCE FROM N.A.
RP STRAIN-FRANCE 1;
RC MEDLINE=20050669; PubMed=10581279;
RA Schmid K.J., Nigro L., Aquadro C.F., Tautz D.;
RT "Large number of replacement polymorphisms in rapidly evolving genes
of Drosophila. Implications for genome-wide surveys of DNA
```

```
polymorphism.";
RL Genetics 153:1717-1729(1999).
DR EMBL; AF161729; AAD45733.1;
DR FlyBase; FBgn0022349; CG1910.
FT NON_TER 310 310
SQ SEQUENCE 310 AA; 32336 MW; 10CF3F4A40A1E13A CRC64;

Query Match 9.9%; Score 77; DB 5; Length 310;
Best Local Similarity 22.2%; Pred. No. 3.4;
Matches 24; Conservative 19; Mismatches 65; Indels 0; Gaps 0;

QY 24 APEKALRNEERAIDELKKQAIEDKEATTAIEAASSDALEALADQTDALQSEEAHVVDNA 83
   I : : I : : I : : I : : I : : I : : I : : I : : I : : I : : I : :
Db 148 APETEAATAKTTESNDVVVEATGKATVLEVPAAAPKAEKAEKAEKAEKAEKAEKAEKAE 207
   I : : I : : I : : I : : I : : I : : I : : I : : I : : I : : I : :
QY 84 NAASDALEALADQTDALQSEEAHVVDNAASDAWEKAATPIALDVKK 131
   I : : I : : I : : I : : I : : I : : I : : I : : I : : I : : I : :
Db 208 KEAASCEEPKSVVDVSAEPKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAE 255
   I : : I : : I : : I : : I : : I : : I : : I : : I : : I : : I : :

RESULT 36
Q91255 PRELIMINARY; PRT; 1110 AA.
ID Q91255
AC Q91255;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE NF-180.
OS Petromyzon marinus (Sea lamprey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Hyperoartia;
OC Petromyzontiformes; Petromyzontidae; Petromyzon.
ON NCBI_TaxID=7757;
RX SEQUENCE FROM N.A.
RP TISSUE=CNS;
RC MEDLINE=95287814; PubMed=7770000;
RA Jacobs A.J., Kamholz J., Selzer M.E.;
RT "The single lamprey neurofilament subunit (NF-180) lacks
RT multiphosphorylation repeats and is expressed selectively in
RT projection neurons.";
RL Brain Res. Mol. Brain Res. 29:43-52(1995).
DR EMBL; U19361; AAA80106.1;
DR InterPro; IPR001664; IF.
DR InterPro; IPR000533; Tropomyosin.
DR Pfam; PF00038; filament.1.
DR PRINTS; PR00194; TROPOMYOSIN.
DR PROSITE; PS00226; IF; UNKNOWN.1.
SQ SEQUENCE 1110 AA; 123818 MW; 6558DA73DAF6974C CRC64;

Query Match 9.9%; Score 77; DB 13; Length 1110;
Best Local Similarity 23.1%; Pred. No. 13;
Matches 21; Conservative 24; Mismatches 46; Indels 0; Gaps 0;

QY 55 EAASSDALEALADQTDALQSEEAHVVDNAASDALEALADQTDALQSEEAHVVDNA 114
   I : : I : : I : : I : : I : : I : : I : : I : : I : : I : : I : :
Db 706 EAEEVEEVEEVTKSDEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAE 765
   I : : I : : I : : I : : I : : I : : I : : I : : I : : I : : I : :
QY 115 SDAMEKAATPIALDVKKTKDTFVVKKEERQ 145
   I : : I : : I : : I : : I : : I : : I : : I : : I : : I : : I : :
Db 766 EEAEKAEASDDEKPEEEVKESEAPVAPPAKK 796
   I : : I : : I : : I : : I : : I : : I : : I : : I : : I : : I : :

RESULT 37
Q21022 PRELIMINARY; PRT; 1133 AA.
ID Q21022
AC Q21022; Q21071;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE F59A2.6 PROTEIN.
GN F59A2.6.
OS Caenorhabditis elegans.
```



BEST LOCAL SIMILARITY 27.38; Pred. NO. 3.9;  
Matches 21; Conservative 13; Mismatches 43; Indels 0; Gaps 0;



```

RP SEQUENCE FROM N.A.
RX STRAIN=PADUA/ITALY 15, AND PADUA/ITALY 14;
RX MEDLINE=20050669; PubMed=10581279;
RA Schmid K.J., Nigro L., Aquadro C.F., Tautz D.;
RT "Large number of replacement polymorphisms in rapidly evolving genes
RT of Drosophila. Implications for genome-wide surveys of DNA
RT polymorphism.";
RL Genetics 153:1717-1729(1999).
DR EMBL; AF161738; AAD45743.1; -
DR EMBL; AF161738; AAD45742.1; -
DR FlyBase; FBgn0022349; CG1910.
FT NON_TER
FT NON_TER
SQ SEQUENCE 310 AA; 32392 MW; F7657BECDA2F81E CRC64;

Query Match 10.8%; Score 84; DB 5; Length 310;
Best Local Similarity 23.1%; Pred. No. 0.61;
Matches 25; Conservative 20; Mismatches 63; Indels 0; Gaps 0;

OY 24 APEKALNEERAIDELKQAEKATTAIEAASDALEALADQTDALQSEEAHVVKAD 83
DB 148 APEEAANKTSSNDVVEVATEGKTVLEVPAAEPKEASTVESAEELTETSTVVVTEP 207

OY 84 NAASDALEALADQTDALQSEEAHVQSDNAASDAWEKAATPIALDYK 131
DB 208 KEAASSEEPSKVVDSPAEPKEASTVESATPPIDVSTAPASNDVSK 255

RESULT 22
OY9VU3 PRELIMINARY; PRT; 489 AA.
AC OY9VU3; Q9VW86;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE ANON-EST:FEI3A3 OR ANONIA3.
GN CG1910 OR ANON-EST:FEI3A3 OR ANONIA3.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananthanides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abil J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cantarel D., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacieb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

```

```

RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE OF 180-489 FROM N.A.
RX STRAIN-VARIOUS STRAINS;
RX MEDLINE=20050669; PubMed=10581279;
RA Schmid K.J., Nigro L., Aquadro C.F., Tautz D.;
RT "Large number of replacement polymorphisms in rapidly evolving genes
RT of Drosophila. Implications for genome-wide surveys of DNA
RT polymorphism.";
RL Genetics 153:1717-1729(1999).
DR EMBL; AF003779; AAF57190.1; -
DR EMBL; AF161747; AAD45751.1; -
DR EMBL; AF161723; AAD45727.1; -
DR EMBL; AF161727; AAD45731.1; -
DR EMBL; AF161732; AAD45736.1; -
DR EMBL; AF161737; AAD45741.1; -
DR EMBL; AF161740; AAD45744.1; -
DR EMBL; AF161741; AAD45745.1; -
DR EMBL; AF161742; AAD45746.1; -
DR EMBL; AF161745; AAD45749.1; -
DR EMBL; AF161746; AAD45750.1; -
DR FlyBase; FBgn0022349; CG1910.
SQ SEQUENCE 489 AA; 51587 MW; 85BBFA05060A9FD2 CRC64;

Query Match 10.8%; Score 84; DB 5; Length 489;
Best Local Similarity 23.1%; Pred. No. 1;
Matches 25; Conservative 20; Mismatches 63; Indels 0; Gaps 0;

OY 24 APEKALNEERAIDELKQAEKATTAIEAASDALEALADQTDALQSEEAHVVKAD 83
DB 327 APEEAANKTSSNDVVEVATEGKTVLEVPAAEPKEASTVESAEELTETSTVVVTEP 386

OY 84 NAASDALEALADQTDALQSEEAHVQSDNAASDAWEKAATPIALDYK 131
DB 387 KEAASSEEPSKVVDSPAEPKEASTVESATPPIDVSTAPASNDVSK 434

RESULT 23
OY9G0H8 PRELIMINARY; PRT; 540 AA.
AC OY9G0H8;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE GP7.
OS Roseophaga SI01.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Podoviridae;
OC T7-like phages.
OX NCBI_TaxID=136084;
RN [1]
RP SEQUENCE FROM N.A.
RA Rohwer F.L., Segall A.M., Steward G., Seguritan V., Breitbart M.,
RA Wolven F., Azam F.;
RT "The complete genomic sequence of the marine phage Roseophaga SI01
RL Limnol. Oceanogr. 45:408-418(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Rohwer F.L., Segall A.M., Steward G., Seguritan V., Breitbart M.,
RA Wolven F., Azam F.;
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF189021; AAG02590.1; -
DR InterPro; IPR000104; Antifreeze_1.

```





```

RX MEDLINE-20050669; PubMed-10581279;
RA Schmid K.J., Nigro L., Aquadro C.F., Tautz D.;
RT "Large number of replacement polymorphisms in rapidly evolving genes
RT of Drosophila. Implications for genome-wide surveys of DNA
RT polymorphism.";
RL Genetics 153:1717-1729(1999).
DR EMBL; AF161724; AAD45728.1; -.
DR FlyBase; FBgn0022349; CG1910.
FT NON_TER 1
FT NON_TER 310
SQ SEQUENCE 310 AA; 32391 MW; 00EBE2551606CB57 CRC64;

Query Match 10.8%; Score 84; DB 5; Length 310;
Best Local Similarity 23.1%; Pred. No. 0.61;
Matches 25; Conservative 20; Mismatches 63; Indels 0; Gaps 0;

QY 24 APEKALRNEERAIDELKKQAIKEDKATTAIEAASSDALEALADQTDALQSEEAAYVKAD 83
Db 148 APETEAANKTSSNDVVVEATGKETVLEVPAAEPKEAESTVESAEELTETSTVVVTEP 207

QY 84 NAASDALEALADQTDALQSEEAAYVQSDNAASDAWEKAATPIALDYKK 131
Db 208 KEAASSEEPKSVVDSEAPKEAESTNTDESATVPIDVSTAPASNDYSK 255

RESULT 15
Q9U4R4
ID Q9U4R4 PRELIMINARY; PRT; 310 AA.
AC Q9U4R4;
RC STRAIN=AUSTRIA 5;
RX MEDLINE-20050669; PubMed-10581279;
RA Schmid K.J., Nigro L., Aquadro C.F., Tautz D.;
RT "Large number of replacement polymorphisms in rapidly evolving genes
RT of Drosophila. Implications for genome-wide surveys of DNA
RT polymorphism.";
RL Genetics 153:1717-1729(1999).
DR EMBL; AF161726; AAD45730.1; -.
DR FlyBase; FBgn0022349; CG1910.
FT NON_TER 1
FT NON_TER 310
SQ SEQUENCE 310 AA; 32372 MW; 122BE25517DABA90 CRC64;

Query Match 10.8%; Score 84; DB 5; Length 310;
Best Local Similarity 23.1%; Pred. No. 0.61;
Matches 25; Conservative 20; Mismatches 63; Indels 0; Gaps 0;

QY 24 APEKALRNEERAIDELKKQAIKEDKATTAIEAASSDALEALADQTDALQSEEAAYVKAD 83
Db 148 APETEAANKTSSNDVVVEATGKETVLEVPAAEPKEAESTVESAEELTETSTVVVTEP 207

QY 84 NAASDALEALADQTDALQSEEAAYVQSDNAASDAWEKAATPIALDYKK 131
Db 208 KEAASSEEPKSVVDSEAPKEAESTNTDESATVPIDVSTAPASNDYSK 255

RESULT 16
Q9U4R3
ID Q9U4R3 PRELIMINARY; PRT; 310 AA.
AC Q9U4R3;
RC STRAIN=ITALY;
RX MEDLINE-20050669; PubMed-10581279;
RA Schmid K.J., Nigro L., Aquadro C.F., Tautz D.;
RT "Large number of replacement polymorphisms in rapidly evolving genes
RT of Drosophila. Implications for genome-wide surveys of DNA
RT polymorphism.";
RL Genetics 153:1717-1729(1999).
DR EMBL; AF161731; AAD45735.1; -.
DR FlyBase; FBgn0022349; CG1910.
FT NON_TER 1
FT NON_TER 310
SQ SEQUENCE 310 AA; 32306 MW; 0EA033FEDA5D25FE CRC64;

Query Match 10.8%; Score 84; DB 5; Length 310;
Best Local Similarity 23.1%; Pred. No. 0.61;
Matches 25; Conservative 20; Mismatches 63; Indels 0; Gaps 0;

QY 24 APEKALRNEERAIDELKKQAIKEDKATTAIEAASSDALEALADQTDALQSEEAAYVKAD 83
Db 148 APETEAANKTSSNDVVVEATGKETVLEVPAAEPKEAESTVESAEELTETSTVVVTEP 207

QY 84 NAASDALEALADQTDALQSEEAAYVQSDNAASDAWEKAATPIALDYKK 131
Db 208 KEAASSEEPKSVVDSEAPKEAESTNTDESATVPIDVSTAPASNDYSK 255

RESULT 17
Q9U4R0
ID Q9U4R0 PRELIMINARY; PRT; 310 AA.
AC Q9U4R0;
RC STRAIN=ITALY;
RX MEDLINE-20050669; PubMed-10581279;
RA Schmid K.J., Nigro L., Aquadro C.F., Tautz D.;
RT "Large number of replacement polymorphisms in rapidly evolving genes
RT of Drosophila. Implications for genome-wide surveys of DNA
RT polymorphism.";
RL Genetics 153:1717-1729(1999).
DR EMBL; AF161731; AAD45735.1; -.
DR FlyBase; FBgn0022349; CG1910.
FT NON_TER 1
FT NON_TER 310
SQ SEQUENCE 310 AA; 32306 MW; 0EA033FEDA5D25FE CRC64;

Query Match 10.8%; Score 84; DB 5; Length 310;
Best Local Similarity 23.1%; Pred. No. 0.61;
Matches 25; Conservative 20; Mismatches 63; Indels 0; Gaps 0;

QY 24 APEKALRNEERAIDELKKQAIKEDKATTAIEAASSDALEALADQTDALQSEEAAYVKAD 83
Db 148 APETEAANKTSSNDVVVEATGKETVLEVPAAEPKEAESTVESAEELTETSTVVVTEP 207

QY 84 NAASDALEALADQTDALQSEEAAYVQSDNAASDAWEKAATPIALDYKK 131
Db 208 KEAASSEEPKSVVDSEAPKEAESTNTDESATVPIDVSTAPASNDYSK 255
```

```

DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE ANONIA3 (FRAGMENT).
GN CG1910 OR ANONIA3.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CYPRIUS;
RX MEDLINE-20050669; PubMed-10581279;
RA Schmid K.J., Nigro L., Aquadro C.F., Tautz D.;
RT "Large number of replacement polymorphisms in rapidly evolving genes
RT of Drosophila. Implications for genome-wide surveys of DNA
RT polymorphism.";
RL Genetics 153:1717-1729(1999).
DR EMBL; AF161728; AAD45732.1; -.
DR FlyBase; FBgn0022349; CG1910.
FT NON_TER 1
FT NON_TER 310
SQ SEQUENCE 310 AA; 32364 MW; 0CCC02A07AA114A1 CRC64;

Query Match 10.8%; Score 84; DB 5; Length 310;
Best Local Similarity 23.1%; Pred. No. 0.61;
Matches 25; Conservative 20; Mismatches 63; Indels 0; Gaps 0;

QY 24 APEKALRNEERAIDELKKQAIKEDKATTAIEAASSDALEALADQTDALQSEEAAYVKAD 83
Db 148 APETEAANKTSSNDVVVEATGKETVLEVPAAEPKEAESTVESAEELTETSTVVVTEP 207

QY 84 NAASDALEALADQTDALQSEEAAYVQSDNAASDAWEKAATPIALDYKK 131
Db 208 KEAASSEEPKSVVDSEAPKEAESTNTDESATVPIDVSTAPASNDYSK 255

RESULT 17
Q9U4R0
ID Q9U4R0 PRELIMINARY; PRT; 310 AA.
AC Q9U4R0;
RC STRAIN=ITALY;
RX MEDLINE-20050669; PubMed-10581279;
RA Schmid K.J., Nigro L., Aquadro C.F., Tautz D.;
RT "Large number of replacement polymorphisms in rapidly evolving genes
RT of Drosophila. Implications for genome-wide surveys of DNA
RT polymorphism.";
RL Genetics 153:1717-1729(1999).
DR EMBL; AF161731; AAD45735.1; -.
DR FlyBase; FBgn0022349; CG1910.
FT NON_TER 1
FT NON_TER 310
SQ SEQUENCE 310 AA; 32306 MW; 0EA033FEDA5D25FE CRC64;

Query Match 10.8%; Score 84; DB 5; Length 310;
Best Local Similarity 23.1%; Pred. No. 0.61;
Matches 25; Conservative 20; Mismatches 63; Indels 0; Gaps 0;

QY 24 APEKALRNEERAIDELKKQAIKEDKATTAIEAASSDALEALADQTDALQSEEAAYVKAD 83
Db 148 APETEAANKTSSNDVVVEATGKETVLEVPAAEPKEAESTVESAEELTETSTVVVTEP 207
```



[illegible]

```

Qy 121 AA 122
Db 121 LA 122

RESULT 5
Q9X5C5
ID Q9X5C5 PRELIMINARY; PRT; 239 AA.
AC Q9X5C5;
DT 01-NOV-1999 (TRENBLrel. 12, Created)
DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE GRAB (FRAGMENT).
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KTL9;
RX MEDLINE=99269061; PubMed=10336419;
RA Rasmussen M., Muller H.P., Bjorck L.;
RT *Protein GRAB of streptococcus pyogenes regulates proteolysis at the
RT bacterial surface by binding alpha2-macroglobulin.*;
RL J. Biol. Chem. 274:15336-15344(1999).
DR EMBL: AF124402; AAD26341.1; -.
DR InterPro: IPR001899; Gram_pos_anchor.
DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
FT NON_TER 1
FT NON_TER 239
SQ SEQUENCE 239 AA; 24846 MW; 61AC4F6F863AF0F5 CRC64;

Query Match 68.7%; Score 534; DB 2; Length 239;
Best Local Similarity 91.8%; Pred. No. 6.2e-49;
Matches 112; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

Qy 1 VDSPIEQRIIPNGGTLNLLGNAPEKALRNEERAIDELKKQATDEKATTAIEAASD 60
Db 1 VDSPIEQRIIPNGGTLNLLGNAPEKALRNEERAIDELKKQATDEKATTAIEAASD 60

Qy 61 ALEALADQTDALQSEAAVVKADNAASDALEALADQTDALQSEAAVVSQDAAASDAWEK 120
Db 61 ALEALADQADALQSEAAVVSQDAAASDALEALADQTDALQSEAAVVKADNAASDTEA 120

Qy 121 AA 122
Db 121 LA 122

RESULT 6
Q54180
ID Q54180 PRELIMINARY; PRT; 103 AA.
AC Q54180;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE PROTEIN G'.
GN PROTEIN G' GENE.
OS Streptococcus sp. 'group G'.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1320;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=G148;
RX MEDLINE=90226312; PubMed=2183792;
RA Goward C.R., Murphy J.P., Atkinson T., Barstow D.A.;
RT *Expression and purification of a truncated recombinant streptococcal
RT protein G.*;
RL Biochem. J. 267:171-177(1990).
DR EMBL: X53324; CAA37409.1; -.
SQ SEQUENCE 103 AA; 10828 MW; E7AE8BD1C073423C CRC64;

```

```

Query Match 18.3%; Score 142; DB 2; Length 103;
Best Local Similarity 57.9%; Pred. No. 1.3e-07;
Matches 33; Conservative 3; Mismatches 21; Indels 0; Gaps 0;

Qy 1 VDSPIEQRIIPNGGTLNLLGNAPEKALRNEERAIDELKKQATDEKATTAIEA 57
Db 34 VDSPIEDPIIRNGGELTNLLGNSSETTLALRNEESATADLTAAAVADTVAAAAAENA 90

RESULT 7
Q53900
ID Q53900 PRELIMINARY; PRT; 323 AA.
AC Q53900;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE ALBUMIN-BINDING PROTEIN (FRAGMENT).
OS Streptococcus canis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DG12;
RX MEDLINE=92363555; PubMed=1500168;
RA Sjobring U.;
RT *Isolation and molecular characterization of a novel albumin-binding
RT protein from group G streptococci.*;
RL Infect. Immun. 60:3601-3608(1992).
DR EMBL: M95520; AAA26847.1; -.
DR HSSP: Q51911; IGAB.
DR InterPro: IPR002988; GA.
DR InterPro: IPR001899; Gram_pos_anchor.
DR Pfam: PF01468; GA; 2.
DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
FT NON_TER 1
FT NON_TER 323
SQ SEQUENCE 323 AA; 34990 MW; E6903CFDCD5D373E CRC64;

Query Match 16.2%; Score 126; DB 2; Length 323;
Best Local Similarity 42.6%; Pred. No. 2.2e-05;
Matches 26; Conservative 14; Mismatches 21; Indels 0; Gaps 0;

Qy 62 LEALADQTDALQSEAAVVKADNAASDALEALADQTDALQSEAAVVSQDAAASDAWEKA 121
Db 62 LSAQADQIVSAQADNEAITKAEEDSSKAWAAADQANTAKAEADELAKAEKSSDAWEKA 121

Qy 122 A 122
Db 122 A 122

RESULT 8
Q935B1
ID Q935B1 PRELIMINARY; PRT; 850 AA.
AC Q935B1;
DT 01-DEC-2001 (TRENBLrel. 19, Created)
DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE PUTATIVE PHAGE TAIL PROTEIN.
GN HCM2.0051C.
OS Salmonella enterica subsp. enterica serovar Typhi.
OX Plasmid pHC2.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=90370;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,

```

```

FT CHAIN 34 217 GRAB.
SQ SEQUENCE 217 AA; 22836 MW; 79AA8C4FF5F3FA06 CRC64;

Query Match 100.0%; Score 777; DB 16; Length 217;
Best Local Similarity 100.0%; Pred. No. 7.9e-75;
Matches 159; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VDSPIEQPRIIPNGGTLTNLLGNAPEKALRNEERATDELKQKAIEDKEATTAEAASSD 60
Db 34 VDSPIEQPRIIPNGGTLTNLLGNAPEKALRNEERATDELKQKAIEDKEATTAEAASSD 93
Qy 61 ALEALADQTDALQSEEAHVKADNAASDALEALADQTDALQSEEAHVQSDNAASDAWEK 120
Db 94 ALEALADQTDALQSEEAHVKADNAASDALEALADQTDALQSEEAHVQSDNAASDAWEK 153
Qy 121 AATPIALDVKKTKDTPVVKKEERQNVNTLPTTGEENP 159
Db 154 AATPIALDVKKTKDTPVVKKEERQNVNTLPTTGEENP 192

RESULT 2
Q9S6G4 PRELIMINARY; PRT; 156 AA.
AC Q9S6G4;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE GRAB (FRAGMENT).
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KTL3;
RX MEDLINE=99269061; PubMed=10336419;
RA Rasmussen M., Muller H.P., Bjorck L.;
RT "Protein GRAB of streptococcus pyogenes regulates proteolysis at the
bacterial surface by binding alpha2-macroglobulin.";
RL J. Biol. Chem. 274:15336-15344(1999).
DR EMBL: AF124401; AAD26340.1; -.
DR InterPro: IPR001899; Gram_pos_anchor.
DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
FT NON_TER 1
FT NON_TER 156
SQ SEQUENCE 156 AA; 16575 MW; 0C20967E2FB0D866 CRC64;

Query Match 97.8%; Score 760; DB 2; Length 156;
Best Local Similarity 100.0%; Pred. No. 3.6e-73;
Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VDSPIEQPRIIPNGGTLTNLLGNAPEKALRNEERATDELKQKAIEDKEATTAEAASSD 60
Db 1 VDSPIEQPRIIPNGGTLTNLLGNAPEKALRNEERATDELKQKAIEDKEATTAEAASSD 60
Qy 61 ALEALADQTDALQSEEAHVKADNAASDALEALADQTDALQSEEAHVQSDNAASDAWEK 120
Db 61 ALEALADQTDALQSEEAHVKADNAASDALEALADQTDALQSEEAHVQSDNAASDAWEK 120
Qy 121 AATPIALDVKKTKDTPVVKKEERQNVNTLPTTGEENP 156
Db 121 AATPIALDVKKTKDTPVVKKEERQNVNTLPTTGEENP 156

RESULT 3
Q9S6G3 PRELIMINARY; PRT; 156 AA.
AC Q9S6G3;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE GRAB (FRAGMENT).
OS Streptococcus pyogenes.

```

```

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AP1;
RX MEDLINE=99269061; PubMed=10336419;
RA Rasmussen M., Muller H.P., Bjorck L.;
RT "Protein GRAB of streptococcus pyogenes regulates proteolysis at the
bacterial surface by binding alpha2-macroglobulin.";
RL J. Biol. Chem. 274:15336-15344(1999).
DR EMBL: AF124400; AAD26339.1; -.
DR InterPro: IPR001899; Gram_pos_anchor.
DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
FT NON_TER 1
FT NON_TER 156
SQ SEQUENCE 156 AA; 16587 MW; 19FAA98D0599D866 CRC64;

Query Match 97.0%; Score 754; DB 2; Length 156;
Best Local Similarity 99.4%; Pred. No. 1.5e-72;
Matches 155; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 VDSPIEQPRIIPNGGTLTNLLGNAPEKALRNEERATDELKQKAIEDKEATTAEAASSD 60
Db 1 VDSPIEQPRIIPNGGTLTNLLGNAPEKALRNEERATDELKQKAIEDKEATTAEAASSD 60
Qy 61 ALEALADQTDALQSEEAHVKADNAASDALEALADQTDALQSEEAHVQSDNAASDAWEK 120
Db 61 ALEALADQTDALQSEEAHVKADNAASDALEALADQTDALQSEEAHVQSDNAASDAWEK 120
Qy 121 AATPIALDVKKTKDTPVVKKEERQNVNTLPTTGEENP 156
Db 121 AATPIALDVKKTKDTPVVKKEERQNVNTLPTTGEENP 156

RESULT 4
Q9X5C6 PRELIMINARY; PRT; 268 AA.
AC Q9X5C6;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE GRAB (FRAGMENT).
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AP49;
RX MEDLINE=99269061; PubMed=10336419;
RA Rasmussen M., Muller H.P., Bjorck L.;
RT "Protein GRAB of streptococcus pyogenes regulates proteolysis at the
bacterial surface by binding alpha2-macroglobulin.";
RL J. Biol. Chem. 274:15336-15344(1999).
DR EMBL: AF124403; AAD26342.1; -.
DR InterPro: IPR001899; Gram_pos_anchor.
DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
FT NON_TER 1
FT NON_TER 268
SQ SEQUENCE 268 AA; 27744 MW; 1C59239260C0C7E7 CRC64;

Query Match 69.0%; Score 536; DB 2; Length 268;
Best Local Similarity 92.6%; Pred. No. 4.3e-49;
Matches 113; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

Qy 1 VDSPIEQPRIIPNGGTLTNLLGNAPEKALRNEERATDELKQKAIEDKEATTAEAASSD 60
Db 1 VDSPIEQPRIIPNGGTLTNLLGNAPEKALRNEERATDELKQKAIEDKEATTAEAASSD 60
Qy 61 ALEALADQTDALQSEEAHVKADNAASDALEALADQTDALQSEEAHVQSDNAASDAWEK 120
Db 61 ALEALADQTDALQSEEAHVQSDNAASDALEALADQTDALQSEEAHVQSDNAAGDALEA 120

```

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2002, 01:14:06 ; Search time 70.5722 Seconds  
(without alignments)  
389.760 Million cell updates/sec

Title: US-09-847-539A-6  
Perfect score: 777  
Sequence: 1 VDSPIEQPRIIPNGTTLNL.....KKEERQNVNTLPTTGESEN 159

Scoring table: BLOSUM62

Gapop 60.0 , Gapext 60.0

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_19:\*

1: sp\_archaea:\*

2: sp\_bacteria:\*

3: sp\_fungi:\*

4: sp\_human:\*

5: sp\_invertebrate:\*

6: sp\_mammal:\*

7: sp\_mhc:\*

8: sp\_organelle:\*

9: sp\_phage:\*

10: sp\_plant:\*

11: sp\_rodent:\*

12: sp\_virus:\*

13: sp\_vertebrate:\*

14: sp\_unclassified:\*

15: sp\_virus:\*

16: sp\_bacterioph:\*

17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	777	100.0	217	16 Q9X5C4	Q9X5C4 streptococc
2	760	97.8	156	2 Q9S6G4	Q9S6G4 streptococc
3	754	97.0	156	2 Q9S6G3	Q9S6G3 streptococc
4	536	69.0	268	2 Q9X5C6	Q9X5C6 streptococc
5	534	68.7	239	2 Q9X5C5	Q9X5C5 streptococc
6	142	18.3	103	2 Q94180	Q94180 streptococc
7	126	16.2	323	2 Q93900	Q93900 streptococc
8	104	13.4	850	2 Q935B1	Q935B1 salmonella
9	91	11.7	413	2 Q93974	Q93974 streptococc
10	90	11.6	619	5 Q95S18	Q95S18 drosophila
11	90	11.6	2148	5 Q9BJD3	Q9BJD3 physarum po
12	89	11.5	447	2 Q9AKA4	Q9AKA4 streptococc
13	85	10.9	664	2 Q93975	Q93975 streptococc
14	84	10.8	310	5 Q9U4R5	Q9U4R5 drosophila
15	84	10.8	310	5 Q9U4R4	Q9U4R4 drosophila
16	84	10.8	310	5 Q9U4R3	Q9U4R3 drosophila

17 84 10.8 310 5 Q9U4R0  
18 84 10.8 310 5 Q9U4Q9  
19 84 10.8 310 5 Q9U4Q8  
20 84 10.8 310 5 Q9U4Q7  
21 84 10.8 310 5 Q9TVG2  
22 84 10.8 489 5 Q9V9U3  
23 84 10.8 540 9 Q9G0H8  
24 83 10.7 310 5 Q9TW87  
25 81 10.4 310 5 Q9U4Q6  
26 81 10.4 558 2 Q54718  
27 81 10.4 669 2 Q93EM8  
28 81 10.4 1272 2 Q9F553  
29 80 10.3 274 11 Q9CQ47  
30 80 10.3 372 2 Q9WWX1  
31 80 10.3 3072 2 Q939N5  
32 78 10.0 113 3 Q08245  
33 78 10.0 576 2 Q9KJ33  
34 78 10.0 2178 2 Q9KWR3  
35 77 9.9 310 5 Q9U4R2  
36 77 9.9 1110 13 Q91255  
37 77 9.9 1133 5 Q21022  
38 76 9.8 395 2 Q937K4  
39 76 9.8 725 13 Q07718  
40 76 9.8 2314 12 Q91LF6  
41 75 9.7 211 11 Q99JX1  
42 75 9.7 509 10 Q9C8J7  
43 75 9.7 791 13 Q9DGL1  
44 75 9.7 969 5 Q9NDI9  
45 75 9.7 2478 2 Q9LCH2

## ALIGNMENTS

### RESULT 1

Q9X5C4 PRELIMINARY; PRT; 217 AA.  
AC Q9X5C4;  
DT 01-NOV-1999 (TRENBLrel. 12, Created)  
DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)  
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)  
DE GRAB PRECURSOR (PROTEIN GRAB) (PROTEIN G-RELATED ALPHA 2N-BINDING PROTEIN).  
GN GRAB OR SPY1357.  
OS Streptococcus pyogenes.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1314;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC700294;  
RA Rasmussen M., Muller H.P., Bjorck L.;  
RT "Protein GRAB of Streptococcus pyogenes regulates proteolysis at the bacterial surface by binding alpha2-macroglobulin.";  
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SF370 / ATCC 700294 / SEROTYPE M1;  
RX MEDLINE=21192684; PubMed=11296296;  
RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K., Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P., Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J., Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;  
RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes.";  
RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).  
DR EMBL: AF124399; AAD26338.1;  
DR EMBL: AE006573; AAK34185.1;  
DR InterPro: IPR001899; Gram\_pos\_anchor.  
DR Pfam: PF00746; Gram\_pos\_anchor; 1.  
DR PRINTS: PS00015; GPOSANCHOR.  
DR PROSITE: PS00343; GRAM\_POS\_ANCHORING; UNKNOWN\_1.  
KW Complete proteome; Signal; Transmembrane.  
FT SIGNAL 1 33 POTENTIAL.

THIS PAGE BLANK (USPTO)

```

; APPLICANT: SHARMA, Lawrence Rajendra
; TITLE OF INVENTION: PEPTIDES REPRESENTING EPITOPIC SITES FOR
; TITLE OF INVENTION: BACTERIAL AND VIRAL MENINGITIS CAUSING AGENTS AND THEIR
; TITLE OF INVENTION: CNS CARRIER, ANTIBODIES THEREOF, AND USES THEREOF
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/482,847
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/127,499
; FILING DATE: 28-SEP-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 51916/104/INBI
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 695 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
US-08-482-847-23
```

```

Query Match 11.6%; Score 90; DB 1; Length 695;
Best Local Similarity 24.2%; Pred. No. 0.89;
Matches 37; Conservative 23; Mismatches 61; Indels 32; Gaps 4;

QY 24 APEKLAIRNEERA-IDELKKQAIEDKEATTAIEAASSDALEALADQTDALQSEEAAY--- 79
Db 212 APELTKKLEAKLEAEAKKATEAKQKVDAAEVAPQAKIAELENVQVHRLEQELKEIDES 271
QY 80 -----VKADNNAASDALEALADQTDALQSEEAAYVQSONAA----- 114
Db 272 ESEDYAKEGFRAPLQSKLDKAKKLSKLELSDKIDELDAEIAKLEDLKAAEENNNVED 331
QY 115 --SDAWEKAATPIALDVKKTK-DTKPVVKKER 144
Db 332 YFKEGLEKTIARAKKAELEKTEADLKKAVNEPEK 364
```

Search completed: October 13, 2002, 04:50:06  
Job time : 28.6378 secs



APPLICATION NUMBER: US/08/247,491A  
FILING DATE: 23-JUN-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: KOWALSKI, Thomas J.  
REGISTRATION NUMBER: 32,147  
REFERENCE/DOCKET NUMBER: 454312-2041  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-588-0800  
TELEFAX: 212-588-0500  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 648 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-247-491A-2

Query Match 11.6%; Score 90; DB 2; Length 648;  
Best Local Similarity 24.2%; Pred. No. 0.81;  
Matches 37; Conservative 23; Mismatches 61; Indels 32; Gaps 4;  
QY 24 APEKALRNEERA-IDELKKQAIEDKEATTATEAASSDALEALADQTDALQSEEA--- 79  
DB 170 APELTKKLEAKLEAEKATEAKQKVDAAEVAPQAKIAELENOVHRLQELKEIDES 229  
QY 80 -----VKADNAASDALEALADQTDALQSEEAEEVQSDNAA----- 114  
DB 230 ESEDYAKEGFRAPLQSKLDKAKKLSKLEELSDKIDELDAETAKLELDQKAAEENNVED 289  
QY 115 --SDAWEKAATPIALDVKKTK-DTRKPVVKKKEER 144  
DB 290 YFKEGLEKTIAAKKAEELEKTEADLKKAVNEPEK 322

RESULT 38  
US-08-446-201-3  
Sequence 3; Application US/08446201B  
Patent No. 6042838  
GENERAL INFORMATION:  
APPLICANT: BRILES, David E.  
TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS FOR MUCOSAL ADMINISTRATION OF  
FILE REFERENCE: 454312-2018  
CURRENT APPLICATION NUMBER: US/08/446,201B  
EARLIER FILING DATE: 1995-05-19  
EARLIER APPLICATION NUMBER: 08/312,949  
EARLIER FILING DATE: 1994-09-30  
EARLIER APPLICATION NUMBER: 08/246,636  
EARLIER FILING DATE: 1994-05-20  
EARLIER APPLICATION NUMBER: 08/048,896  
EARLIER FILING DATE: 1993-04-20  
EARLIER APPLICATION NUMBER: 07/835,698  
EARLIER FILING DATE: 1992-02-12  
EARLIER APPLICATION NUMBER: 07/656,773  
EARLIER FILING DATE: 1991-02-15  
NUMBER OF SEQ ID NOS: 4  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 3  
LENGTH: 648  
TYPE: PRT  
ORGANISM: Streptococcus pneumoniae  
US-08-446-201-3

Query Match 11.6%; Score 90; DB 3; Length 648;  
Best Local Similarity 24.2%; Pred. No. 0.81;  
Matches 37; Conservative 23; Mismatches 61; Indels 32; Gaps 4;  
QY 24 APEKALRNEERA-IDELKKQAIEDKEATTATEAASSDALEALADQTDALQSEEA--- 79  
DB 170 APELTKKLEAKLEAEKATEAKQKVDAAEVAPQAKIAELENOVHRLQELKEIDES 229

QY 80 -----VKADNAASDALEALADQTDALQSEEAEEVQSDNAA----- 114  
DB 230 ESEDYAKEGFRAPLQSKLDKAKKLSKLEELSDKIDELDAETAKLELDQKAAEENNVED 289  
QY 115 --SDAWEKAATPIALDVKKTK-DTRKPVVKKKEER 144  
DB 290 YFKEGLEKTIAAKKAEELEKTEADLKKAVNEPEK 322  
RESULT 39  
US-08-127-499A-23  
Sequence 23; Application US/08127499A  
Patent No. 5510264  
GENERAL INFORMATION:  
APPLICANT: VAN ALSTYNE, Diane  
TITLE OF INVENTION: ANTIBODIES WHICH BIND MENINGITIS RELATED  
TITLE OF INVENTION: HOMOLOGOUS ANTIGENIC SEQUENCES  
NUMBER OF SEQUENCES: 40  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/127,499A  
FILING DATE: 28-SEP-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 51916/102/INBI  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 23:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 695 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: unknown  
US-08-127-499A-23

Query Match 11.6%; Score 90; DB 1; Length 695;  
Best Local Similarity 24.2%; Pred. No. 0.89;  
Matches 37; Conservative 23; Mismatches 61; Indels 32; Gaps 4;  
QY 24 APEKALRNEERA-IDELKKQAIEDKEATTATEAASSDALEALADQTDALQSEEA--- 79  
DB 212 APELTKKLEAKLEAEKATEAKQKVDAAEVAPQAKIAELENOVHRLQELKEIDES 271  
QY 80 -----VKADNAASDALEALADQTDALQSEEAEEVQSDNAA----- 114  
DB 272 ESEDYAKEGFRAPLQSKLDKAKKLSKLEELSDKIDELDAETAKLELDQKAAEENNVED 331  
QY 115 --SDAWEKAATPIALDVKKTK-DTRKPVVKKKEER 144  
DB 332 YFKEGLEKTIAAKKAEELEKTEADLKKAVNEPEK 364

RESULT 40  
US-08-482-847-23  
Sequence 23; Application US/08482847  
Patent No. 5556757  
GENERAL INFORMATION:  
APPLICANT: VAN ALSTYNE, Diane



TELEX: LUKPAT WASHINGTON  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 648 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-072-070-2

Query Match 11.6% Score 90; DB 1; Length 648;  
Best Local Similarity 24.2%; Pred. No. 0.81;  
Matches 37; Conservative 23; Mismatches 61; Indels 32; Gaps 4;  
QY 24 APEKALRNEERA-IDELKKAIEDKEATTAEAASSDALEALADQTDALQSEAAV--- 79  
Db 170 APELTKKLEEAKEAEKATEAKQVDAEEVAPQAKIAELENOVHRLQEQLKEIDES 229  
QY 80 -----VKADNNAASDALEALADQTDALQSEAAEVQSDNAA----- 114  
Db 230 ESEDYAKGFRAPLQSKLDKAKKLSLELSKDIDELDAEIAKLEDLQKAAEENNVNVED 289  
QY 115 --SDAWKAAATPIALDVKKTK-DTKPVVKKER 144  
Db 290 YFKEGLEKTTIAAKKAELEKTEADLKAVNEPEK 322

RESULT 33  
US-08-469-434-2  
Sequence 2, Application US/08469434  
Patent No. 5753463  
GENERAL INFORMATION:  
APPLICANT: Briles, David E  
APPLICANT: Yother, Janet L  
APPLICANT: McDaniel, Larry S  
TITLE OF INVENTION: STRUCTURAL GENE OF PNEUMOCOCCAL  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Shoemaker and Mattare, Ltd  
STREET: Suite 1203, 2001 Jefferson Davis Highway  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202-0286  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/469,434  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/072,065  
FILING DATE: 03 JUNE 1993  
APPLICATION NUMBER: US/07/835,698  
FILING DATE: 12-FEB-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/656,773  
FILING DATE: 15-FEB-1991  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 415-0810  
TELEFAX: (703) 521-0378  
TELEX: LUKPAT WASHINGTON  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 648 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-469-434-2

Query Match 11.6% Score 90; DB 1; Length 648;  
Best Local Similarity 24.2%; Pred. No. 0.81;  
Matches 37; Conservative 23; Mismatches 61; Indels 32; Gaps 4;  
QY 24 APEKALRNEERA-IDELKKAIEDKEATTAEAASSDALEALADQTDALQSEAAV--- 79  
Db 170 APELTKKLEEAKEAEKATEAKQVDAEEVAPQAKIAELENOVHRLQEQLKEIDES 229  
QY 80 -----VKADNNAASDALEALADQTDALQSEAAEVQSDNAA----- 114  
Db 230 ESEDYAKGFRAPLQSKLDKAKKLSLELSKDIDELDAEIAKLEDLQKAAEENNVNVED 289  
QY 115 --SDAWKAAATPIALDVKKTK-DTKPVVKKER 144  
Db 290 YFKEGLEKTTIAAKKAELEKTEADLKAVNEPEK 322  
RESULT 34  
US-08-214-222-2  
Sequence 2, Application US/08214222  
Patent No. 5804193  
GENERAL INFORMATION:  
APPLICANT: Briles, David E  
APPLICANT: Yother, Janet L  
TITLE OF INVENTION: STRUCTURAL GENE OF PNEUMOCOCCAL  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Shoemaker and Mattare, Ltd  
STREET: Suite 1203, 2001 Jefferson Davis Highway  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202-0286  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/214,222  
FILING DATE: 17-MAR-1994  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/835,698  
FILING DATE: 12-FEB-1992  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 415-0810  
TELEFAX: (703) 521-0378  
TELEX: LUKPAT WASHINGTON  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 648 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-214-222-2  
Query Match 11.6% Score 90; DB 1; Length 648;  
Best Local Similarity 24.2%; Pred. No. 0.81;  
Matches 37; Conservative 23; Mismatches 61; Indels 32; Gaps 4;  
QY 24 APEKALRNEERA-IDELKKAIEDKEATTAEAASSDALEALADQTDALQSEAAV--- 79  
Db 170 APELTKKLEEAKEAEKATEAKQVDAEEVAPQAKIAELENOVHRLQEQLKEIDES 229  
QY 80 -----VKADNNAASDALEALADQTDALQSEAAEVQSDNAA----- 114  
Db 230 ESEDYAKGFRAPLQSKLDKAKKLSLELSKDIDELDAEIAKLEDLQKAAEENNVNVED 289  
QY 115 --SDAWKAAATPIALDVKKTK-DTKPVVKKER 144  
Db 290 YFKEGLEKTTIAAKKAELEKTEADLKAVNEPEK 322





[illegible]









```

RESULT 15
US-08-591-079-8
: Sequence 8, Application US/08591079
: Patent No. 5972899
: GENERAL INFORMATION:
: APPLICANT: Zychlinsky, Arturo
: APPLICANT: Chen, Yajing
: TITLE OF INVENTION: Apoptosis Induced by Shigella Ipab
: NUMBER OF SEQUENCES: 10
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: MORRISON & FOERSTER
: STREET: 2000 Pennsylvania Avenue, NW
: CITY: Washington
: STATE: DC
: COUNTRY: USA
: ZIP: 20006-1812
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: FILING DATE: US/08/591,079
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Livnat, Shmuel
: REGISTRATION NUMBER: 33,949
: REFERENCE/DOCKET NUMBER: 15661-20017.00
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202) 887-1500
: TELEFAX: (202) 887-0764
: TELEX: 90-4030 MRSNFOERSWSH
: INFORMATION FOR SEQ ID NO: 8:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 593 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
US-08-591-079-8
Query Match 11.7%; Score 91; DB 2; Length 593;
Best Local Similarity 28.7%; Pred. No. 0.58;
Matches 41; Conservative 20; Mismatches 70; Indels 12; Gaps 5;

QY 15 GTLTNLGNA-----PEKALNRNEARAIDELKKAIE~DKEATTATIEAA--SSDALEALA 66
| | | | | : : : : : | : | : | : | : | : | : |
Db 90 GKLTLLGDVSLQSLESRLAV--WQAMIESQKENGIVSKSFQTFALGEAQEATDLYEASI 147
| | | | | : : : : : | : | : | : | : | : | : |
QY 67 DQTDALQSEAAVVKADNAASDALEALADQTDALQSEAEVVSQSDNNAASDAWEKAATPIA 126
| | : | : | : | : | : | : | : | : | : | : | : |
Db 148 KKTDTAKSVYDAATFKULTQGNKQLSLDPADPGYAQAQAAVEQAQKGEATEA--REALDKA 205
| | | | | : | : | : | : | : | : | : | : | : |
QY 127 LDVKKTKDTKPVVKKEERQNVNT 149
| | | | | : | : | : | : | : | : | : | : | : |
Db 206 TDAIVKAGTDAKAKAEKADNILT 228
| | | | | : | : | : | : | : | : | : | : | : |

RESULT 16
US-08-591-079-10
: Sequence 10, Application US/08591079
: Patent No. 5972899
: GENERAL INFORMATION:
: APPLICANT: Zychlinsky, Arturo
: APPLICANT: Chen, Yajing
: TITLE OF INVENTION: Apoptosis Induced by Shigella Ipab
: NUMBER OF SEQUENCES: 10
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: MORRISON & FOERSTER
: STREET: 2000 Pennsylvania Avenue, NW
: CITY: Washington
: STATE: DC

```

```
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 472 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-115-746-10

Query Match 11.9%; Score 92.5; DB 4; Length 472;
Best Local Similarity 22.2%; Pred. No. 0.3; Mismatches 36; Indels 11; Gaps 4;
Matches 34; Conservative 36; Mismatches 36; Indels 11; Gaps 4;

Qy 3 SPIEQRIIPNGTTLNLGNAPE-----KLALRNEERAIDELKQAIEDKQATTAIE 55
Db 219 SDLIEGRGIPPCRNSTKVAEAKQAAEATKVAEAEKQAAEATKVAEAEKQAAEATK 278
Qy 56 AASSDALEALADQTDALQSEAAVVKADNAASDALEALADQTDALQSEAAVVKADNAAS 115
Db 279 VAEAEKQAAEATKVAEAEKQAAEATKVAEAEKQAAEATKVAEAEKQAAEATKVAE 337
Qy 116 DAWKA--ATPIA-LDVKTKTRPVVKKERQ 145
Db 338 AEKQAAEATKVAEAEKQAAEATKVAEAEKQK 370

RESULT 13
; Sequence 2, Application US/08446137B
; Patent No. 6162903
; GENERAL INFORMATION:
; APPLICANT: Trowern, Angus R.
; APPLICANT: Atkinson, Anthony
; APPLICANT: Murphy, Jonathan P.
; APPLICANT: Laurence, Oliver S.
; APPLICANT: Duggieby, Clive J.
; TITLE OF INVENTION: IMMUNOGLOBULIN BINDING PROTEINS DERIVED
; FROM L PROTEIN AND THEIR USES
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/446.137B
; FILING DATE: 22-MAY-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: McMasters, David D.
; REGISTRATION NUMBER: 33,963
; REFERENCE/DOCKET NUMBER: 100084.406
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1027 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-446-137B-2

Query Match 11.9%; Score 92.5; DB 4; Length 1027;
Best Local Similarity 26.8%; Pred. No. 0.86;

; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 472 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-115-746-10

Query Match 11.8%; Score 91.5; DB 2; Length 835;
Best Local Similarity 23.3%; Pred. No. 0.81;
Matches 37; Conservative 30; Mismatches 73; Indels 19; Gaps 4;

Qy 16 TLTNLLGNA---PEKLALRNEERAIDELKQAIEDKQATTAIEAASSDALEALADQTDAL 72
Db 296 TWTALLSRSLQTEKLTVKSQ-ALQE-RDVAIEKQESRVLEQVSAQLECKGQTEQL 353
Qy 73 QSEBAVVKADNAASDALEALADQTDALQSE-----AEVQSDNNAASDAW 118
Db 354 ELENRLATDLRAQLQILANMDSQLKQTHCAQDLAMKDELFCOLTOSNEEQAAQW 413
Qy 119 EKAATPIALDVKKTKTPVVKKEERQNVNLTPTTGES 157
Db 414 QREEMALKHMOAELOQOOQAVLAKEVRLDKETLEFADQEN 452
```

US-08-795-475-6  
; Sequence 6, Application US/08795475  
; Patent No. 5965390  
; GENERAL INFORMATION:  
; APPLICANT: Bivrick, Lars  
; APPLICANT: Sivbring, Ulf  
; TITLE OF INVENTION: PROTEIN L AND HYBRID PROTEINS THEREOF  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED and BERRY LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104-7092  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/795,475  
; FILING DATE: 11-FEB-1997  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: McMasters, David D.  
; REGISTRATION NUMBER: 33,963  
; REFERENCE/DOCKET NUMBER: 100084.402D1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-4900  
; TELEFAX: (206) 682-6031  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 443 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-795-475-6

Query Match 12.0%; Score 93; DB 2; Length 443;  
Best Local Similarity 25.0%; Pred. No. 0.25;  
Matches 44; Conservative 27; Mismatches 47; Indels 58; Gaps 8;  
QY 21 LGNAPEKALNRNEERAIDELKQAIE-----DKEATTATEAASD--ALEAL--A 66  
Db 264 LANLTAEIDKVKKEKQISDASRQLRRDLDSREAKKQVEKALEANSKUALEKLNKL 323  
QY 67 DQTDALQSEEAAYVVKADNAASDALEALADQTDALQSEEAAYVQSDNAASDAWEKAATPIA 126  
Db 324 EESKLUKEKEAELOAKLEAE--AKALKEQL-AKQAEELAKLRAGKASD----- 369  
QY 127 LDVKKTKDKTP-----VVKKEERQHWNTLPTTGEESNP 159  
Db 370 ---SQTPTDKPGNKAVPGKGQAPQAGTKPNQNKAPMKETKRQ----LPSTGETANP 418  
RESULT 11  
US-08-216-894-10  
; Sequence 10, Application US/08216894  
; Patent No. 5876734  
; GENERAL INFORMATION:  
; APPLICANT: Kirchhoff, Louis V.  
; APPLICANT: Otsu, Keiko  
; TITLE OF INVENTION: POLYPEPTIDES FOR DIAGNOSING INFECTION  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 3000 K Street, N.W., Suite 500  
; CITY: Washington, D.C.  
; COUNTRY: USA  
; ZIP: 20007-5109  
; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/216,894  
; FILING DATE: 24-MAR-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BENT, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 85326/102/DRLO  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202)672-5300  
; TELEFAX: (202)672-5399  
; TELEX: 904136  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 472 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-216-894-10  
Query Match 11.9%; Score 92.5; DB 2; Length 472;  
Best Local Similarity 22.2%; Pred. No. 0.3;  
Matches 34; Conservative 36; Mismatches 72; Indels 11; Gaps 4;  
QY 3 SPIEQPRIIPNGTGLTNLGNAP-----KLALRNEERAIDELKKAIEDKEATTAIE 55  
Db 219 SDLTEGRGIPPGCRNSTKVAEAEKQKAAEATKVAEAEKQRAAEATKVAEAEKQKAAEATK 278  
QY 56 AASSDALEALADQTDALQSEEAAYVVKADNAASDALEALADQTDALQSEEAAYVQSDNAAS 115  
Db 279 VAEAEKQRA-AEATKVAEAEKQKAAEATKVGDEKQKAAEATKVAEAEKQKAAEATKVAE 337  
QY 116 DAWEKA--ATPIA-LDVKKTKDKTPVVKKEERQ 145  
Db 338 AEKQKAAEATKVAEAEKQKAAEATKVAEAEKQK 370  
RESULT 12  
US-09-115-746-10  
; Sequence 10, Application US/09115746  
; Patent No. 6228601  
; GENERAL INFORMATION:  
; APPLICANT: Kirchhoff, Louis V.  
; APPLICANT: Otsu, Keiko  
; TITLE OF INVENTION: POLYPEPTIDES FOR DIAGNOSING INFECTION  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 3000 K Street, N.W., Suite 500  
; CITY: Washington, D.C.  
; COUNTRY: USA  
; ZIP: 20007-5109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/115,746  
; FILING DATE:  
; PRIOR APPLICATION NUMBER: 08/216,894  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BENT, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 85326/102/DRLO  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202)672-5300

```
; COUNTRY: U.S.A.
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/005.069
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/630,822
; FILING DATE: 11-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: CONNELL, GARY J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2618-17-C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 62:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 900 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-005-069-62
```

Query Match 13.0%; Score 101; DB 2; Length 900;

Best Local Similarity 25.3%; Pred. No. 0.11; Mismatches 34; Conservative 41; Indels 22; Gaps 5;

```
Qy 17 LTNLLGNAPEKLAIRNEERAIDELKKQAIEDKEATTATAEAASS---DALEALADOTDALQ 73
Db 492 LEKQINDLQEK--LAGELEHNGKLLKQAVELRVASQATEQLNNELOETWQGLQTORDALQ 549
Qy 74 SEEAIVV-----KADNAASD-----ALEALADOTDALQSEAEVVSQSDNAASD--- 116
Db 550 QEVASIQGLKLSQERSRSQASDMQIELEAKLQALHLELHVRCNCEKVTQDNRLRLERIS 609
Qy 117 AWEKAATPIALDVKTKTPVVKKEERQNVNTLPTTGEESN 158
Db 610 TLEKECASLELEKATQNKYEQEVKAHRETEKSRSLVSKEEAN 651
```

## RESULT 8

```
US-09-171-517B-11
; Sequence 11, Application US/09171517B
; Patent No. 6355254
; GENERAL INFORMATION:
; APPLICANT: Finlay, B. Brett
; APPLICANT: Kenny, Brendan
; APPLICANT: Stein, Markus
; TITLE OF INVENTION: PATHOGENIC ESCHERICHIA COLI ASSOCIATED PROTEIN
; FILE REFERENCE: 07422/019001
; CURRENT APPLICATION NUMBER: US/09/171.517B
; CURRENT FILING DATE: 1998-10-20
; PRIOR APPLICATION NUMBER: PCT/CA97/00265
; PRIOR FILING DATE: 1997-04-23
; PRIOR APPLICATION NUMBER: 60/015,999
; PRIOR FILING DATE: 1996-04-23
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: Patentn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 312
; TYPE: PPT
; ORGANISM: Escherichia coli
US-09-171-517B-11
```

Query Match 12.3%; Score 95.5; DB 4; Length 312;

Best Local Similarity 27.6%; Pred. No. 0.09; Mismatches 20; Conservative 43; Indels 23; Gaps 6;

```
Qy 14 GGTLTNLLGNAPEKLAIRNEERAIDEL--KKQAIEDKEATTATAEAASS---DALEALAD- 67
Db 102 GGLISSALGILGSPAAMNNAKGAIEAKASSASSASSKAAGAASEVANKALVKATESVADV 161
Qy 68 ---QTDALQSEEAIVVKADNAASDALEALADOTDALQSEEAIVVSQSDNAASDAWEKAA-- 122
Db 162 ABEASSAMQKAMATTTKAAASRGVADVDYAKATFAED-----LADAAEKTSTRI 210
Qy 123 TPIALDVKK-TKDTKPVVKKEERQNVNTLPTTGEES 157
Db 211 NKLLNSVDKLTNTTAFVAVTSLAEGTKTLPTTISES 246
```

## RESULT 9

```
US-08-158-353-2
; Sequence 2, Application US/08158353
; Patent No. 5620862
; GENERAL INFORMATION:
; APPLICANT: Padula, Steven J.
; TITLE OF INVENTION: Methods for Diagnosing Early Lyme
; TITLE OF INVENTION: Disease
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: MA
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/158,353
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Cartoll, Alice O.
; REGISTRATION NUMBER: 33,542
; REFERENCE/DOCKET NUMBER: UCT93-05
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 212 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-158-353-2
```

Query Match 12.0%; Score 93; DB 1; Length 212;

Best Local Similarity 28.0%; Pred. No. 0.094; Mismatches 46; Conservative 30; Mismatches 58; Indels 30; Gaps 8;

```
Qy 14 GGTLTNLLGNAPEK---LALRNEE---RAIDELKKQAIEDKEATTATAEAASSD--ALEA 64
Db 39 GPNLTEISKKITESNAVVAVKVEVETLLASIDEVAKKATGNLIAQNGNLNAGANQNGSLIA 98
Qy 65 -----LADOTDALQSEEAIVVKADNAASDALEALADOTDALQSEEAIVVSQSDNAASD 116
Db 99 GAYVISTLIAEKLGLKNSSELKEKIE---DAKCKNAFTDKLKSSHAELGIANGAATD 154
Qy 117 AWEKAATPIALDVKTKTKTPVVKKE-ER--QNVNTLPTTGEES 157
Db 155 ANAKAA-----ILKTNGTKDKGAQGLEKLFESVKNLSKAAQET 192
```

## RESULT 10



```
Query Match 14.7%; Score 114.5; DB 3; Length 413;
Best Local Similarity 26.2%; Pred. No. 0.0019;
Matches 49; Conservative 15; Mismatches 56; Indels 67; Gaps 6;

Qy 36 AIDELKKQAIED-----KEATTAIE---AASDALEALADQTDALQSEEAHVVK 81
Db 206 AIEILKYGIGYIKLINNGKTAEGVTALKDEILASKPAVIDAPELTPALTYTKLVI-- 263
Qy 82 ADNAASDALEALADQTDALQSEEA-EVVSQSDNAASDAWE----- 119
Db 264 --NGKTLAGEITTRAADAETAEKAFQYANENGVDGVTYDDATKTFVTETVTEPGDA 321
Qy 120 -----KAATPIALDVKKTKDT-----KPVVKKKEERQNVNTLPT 152
Db 322 PTEPKKPEASIPVLPLTPATPIAKDDAKKDDTKKDDTKKEDAKKPEAKKEAKKAATLPT 381
Qy 153 TGEESNP 159
Db 382 TGEESNP 388

RESULT 2
US-09-171-517B-6
; Sequence 6, Application US/09171517B
; Patent No. 6355254
; GENERAL INFORMATION:
; APPLICANT: Finlay, B. Brett
; APPLICANT: Kenny, Brendan
; APPLICANT: Stein, Markus
; TITLE OF INVENTION: PATHOGENIC ESCHERICHIA COLI ASSOCIATED PROTEIN
; FILE REFERENCE: 07422/019001
; CURRENT APPLICATION NUMBER: US/09/171.517B
; CURRENT FILING DATE: 1998-10-20
; PRIOR APPLICATION NUMBER: PCT/CA97/00265
; PRIOR FILING DATE: 1997-04-23
; PRIOR APPLICATION NUMBER: 60/015,999
; PRIOR FILING DATE: 1996-04-23
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 314
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-171-517B-6

Query Match 13.6%; Score 105.5; DB 4; Length 314;
Best Local Similarity 28.7%; Pred. No. 0.0097;
Matches 45; Conservative 23; Mismatches 68; Indels 21; Gaps 6;

Qy 14 GGTLTNLGNAPKALRNEERAIDELKKQA--IEDKEATT---AIEAASDALEALADQ 68
Db 100 GGAISSVLG-----ILGSFAAINSATKGASDIAQKTASTSSKAIDAASDTATKTLTKA 152

Qy 69 T----DALQSEEAHVVKADNAASDALEALADQTDALQSEEAHVVSQSDNAASDAWEKAATP 124
Db 153 TESVADAVEDASSVMOQAMTTATRAASRTSDVADDIADSAQASQALAEADAQAQKASRA 212

Qy 125 -----IALDVKKTKDTPVVKKEERQNVNTLPTTGEES 157
Db 213 SRFMAAVD-KITGSTPPIAVTSLAEGTKTLPTTVSES 248

RESULT 3
US-09-171-517B-9
; Sequence 9, Application US/09171517B
; Patent No. 6355254
; GENERAL INFORMATION:
; APPLICANT: Finlay, B. Brett
; APPLICANT: Kenny, Brendan
; APPLICANT: Stein, Markus
; TITLE OF INVENTION: PATHOGENIC ESCHERICHIA COLI ASSOCIATED PROTEIN
; FILE REFERENCE: 07422/019001
```

```
; CURRENT APPLICATION NUMBER: US/09/171.517B
; CURRENT FILING DATE: 1998-10-20
; PRIOR APPLICATION NUMBER: PCT/CA97/00265
; PRIOR FILING DATE: 1997-04-23
; PRIOR APPLICATION NUMBER: 60/015,999
; PRIOR FILING DATE: 1996-04-23
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 314
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-171-517B-9

Query Match 13.6%; Score 105.5; DB 4; Length 314;
Best Local Similarity 28.7%; Pred. No. 0.0097;
Matches 45; Conservative 23; Mismatches 68; Indels 21; Gaps 6;

Qy 14 GGTLTNLGNAPKALRNEERAIDELKKQA--TEDKEATT---AIEAASDALEALADQ 68
Db 100 GGAISSVLG-----ILGSFAAINSATKGASDIAQKTASTSSKAIDAASDTATKTLTKA 152

Qy 69 T----DALQSEEAHVVKADNAASDALEALADQTDALQSEEAHVVSQSDNAASDAWEKAATP 124
Db 153 TESVADAVEDASSVMOQAMTTATRAASRTSDVADDIADSAQASQALAEADAQAQKASRA 212

Qy 125 -----IALDVKKTKDTPVVKKEERQNVNTLPTTGEES 157
Db 213 SRFMAAVD-KITGSTPPIAVTSLAEGTKTLPTTVSES 248

RESULT 4
US-09-171-517B-10
; Sequence 10, Application US/09171517B
; Patent No. 6355254
; GENERAL INFORMATION:
; APPLICANT: Finlay, B. Brett
; APPLICANT: Kenny, Brendan
; APPLICANT: Stein, Markus
; TITLE OF INVENTION: PATHOGENIC ESCHERICHIA COLI ASSOCIATED PROTEIN
; FILE REFERENCE: 07422/019001
; CURRENT APPLICATION NUMBER: US/09/171.517B
; CURRENT FILING DATE: 1998-10-20
; PRIOR APPLICATION NUMBER: PCT/CA97/00265
; PRIOR FILING DATE: 1997-04-23
; PRIOR APPLICATION NUMBER: 60/015,999
; PRIOR FILING DATE: 1996-04-23
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 314
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-171-517B-10

Query Match 13.6%; Score 105.5; DB 4; Length 314;
Best Local Similarity 28.7%; Pred. No. 0.0097;
Matches 45; Conservative 23; Mismatches 68; Indels 21; Gaps 6;

Qy 14 GGTLTNLGNAPKALRNEERAIDELKKQA--TEDKEATT---AIEAASDALEALADQ 68
Db 100 GGAISSVLG-----ILGSFAAINSATKGASDIAQKTASTSSKAIDAASDTATKTLTKA 152

Qy 69 T----DALQSEEAHVVKADNAASDALEALADQTDALQSEEAHVVSQSDNAASDAWEKAATP 124
Db 153 TESVADAVEDASSVMOQAMTTATRAASRTSDVADDIADSAQASQALAEADAQAQKASRA 212

Qy 125 -----IALDVKKTKDTPVVKKEERQNVNTLPTTGEES 157
Db 213 SRFMAAVD-KITGSTPPIAVTSLAEGTKTLPTTVSES 248

RESULT 5
```

GenCore version 5.1.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2002, 04:45:42 : Search time 24.6578 Seconds  
(without alignments)  
157.503 Million cell updates/sec

Title: US-09-847-539A-6  
Perfect score: 777  
Sequence: 1 VDSPIEQPRIIPNGOTLTNL.....KKEERQNVNLTPTTGESNP 159

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 2442594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA: \*  
1: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep: \*  
2: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep: \*  
3: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep: \*  
4: /cgn2\_6/ptodata/2/1aa/6B\_COMB.pep: \*  
5: /cgn2\_6/ptodata/2/1aa/PTCUS\_COMB.pep: \*  
6: /cgn2\_6/ptodata/2/1aa/backfiles1.pep: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	114.5	14.7	413	3	US-08-669-408B-10
2	105.5	13.6	314	4	US-09-171-517B-6
3	105.5	13.6	314	4	US-09-171-517B-9
4	105.5	13.6	314	4	US-09-171-517B-10
5	105	13.5	654	3	US-08-669-408B-2
6	101	13.0	900	2	US-08-630-822A-62
7	101	13.0	900	2	US-09-005-069-62
8	95.5	12.3	312	4	US-09-171-517B-11
9	93	12.0	212	1	US-08-158-353-2
10	93	12.0	443	2	US-08-795-475-6
11	92.5	11.9	472	2	US-08-216-894-10
12	92.5	11.9	472	4	US-09-115-746-10
13	92.5	11.9	1027	4	US-08-446-137B-2
14	91.5	11.8	835	2	US-08-968-751-4
15	91	11.7	593	2	US-08-591-079-8
16	91	11.7	593	2	US-08-591-079-10
17	90.5	11.6	516	2	US-08-762-106-8
18	90.5	11.6	516	3	US-08-745-404-2
19	90.5	11.6	516	4	US-09-320-774-8
20	90.5	11.6	527	2	US-08-762-106-9
21	90.5	11.6	527	4	US-09-320-774-9
22	90.5	11.6	552	3	US-08-745-404-3
23	90	11.6	190	1	US-08-106-981-2
24	90	11.6	619	1	US-08-465-746-2
25	90	11.6	619	1	US-08-214-164-2
26	90	11.6	619	2	US-08-467-852A-2
27	90	11.6	619	2	US-08-246-636-2

28	90	11.6	619	2	US-08-247-491A-3	Sequence 3, Appl1
29	90	11.6	619	2	US-08-319-795-2	Sequence 2, Appl1
30	90	11.6	619	2	US-08-468-985-2	Sequence 2, Appl1
31	90	11.6	619	3	US-08-312-949-2	Sequence 2, Appl1
32	90	11.6	648	1	US-08-072-070-2	Sequence 2, Appl1
33	90	11.6	648	1	US-08-469-434-2	Sequence 2, Appl1
34	90	11.6	648	1	US-08-214-222-2	Sequence 2, Appl1
35	90	11.6	648	2	US-08-467-852A-2	Sequence 2, Appl1
36	90	11.6	648	2	US-08-468-718-2	Sequence 2, Appl1
37	90	11.6	648	2	US-08-247-491A-2	Sequence 2, Appl1
38	90	11.6	648	3	US-08-446-201-3	Sequence 3, Appl1
39	90	11.6	695	1	US-08-127-499A-23	Sequence 23, Appl1
40	90	11.6	695	1	US-08-482-847-23	Sequence 23, Appl1
41	89	11.5	289	1	US-08-072-070-4	Sequence 4, Appl1
42	89	11.5	289	1	US-08-469-434-4	Sequence 4, Appl1
43	89	11.5	289	1	US-08-214-222-4	Sequence 4, Appl1
44	89	11.5	289	2	US-08-467-852A-5	Sequence 5, Appl1
45	89	11.5	289	2	US-08-468-718-4	Sequence 4, Appl1

ALIGNMENTS

RESULT 1  
US-08-669-408B-10  
: Sequence 10, Application US/08669408B  
: Patent No. 6100055  
: GENERAL INFORMATION:  
: APPLICANT: GUSS, Bengt  
: APPLICANT: JONSSON, Hans  
: APPLICANT: LINDBERG, Martin  
: APPLICANT: MUELLER, Hans-Peter  
: APPLICANT: RANTAMAKI, Liisa K.  
: TITLE OF INVENTION: METHOD AND MEANS FOR PRODUCING  
: CORRESPONDENCE ADDRESS: 12  
: ADDRESSEE: Foley & Lardner  
: STREET: 3000 K Street, N.W., Suite 500  
: CITY: Washington  
: STATE: D.C.  
: COUNTRY: USA  
: ZIP: 20007-5109  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: PatentIn Release #1.0, Version #1.30  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/08/669,408B  
: FILING DATE: 03-JUL-1996  
: CLASSIFICATION:  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: WO PCT/SE94/00826  
: FILING DATE: 06-SEP-1994  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: SE 9302855-3  
: FILING DATE: 06-SEP-1993  
: ATTORNEY/AGENT INFORMATION:  
: NAME: BENT, Stephen A.  
: REGISTRATION NUMBER: 29,768  
: REFERENCE/DOCKET NUMBER: 61743/102  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: (202)672-5300  
: TELEFAX: (202)672-5399  
: TELEX: 904136  
: INFORMATION FOR SEQ ID NO: 10:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 413 amino acids  
: TYPE: amino acid  
: TOPOLOGY: linear  
: MOLECULE TYPE: protein  
US-08-669-408B-10



**THIS PAGE BLANK (uspto)**

Db 311 AAGGGTACTGAGGCTGCTGAGGATTCTGAGGCTGATGAGGCTGCTGAGGTTTTGAAGGAT 252  
Qy 129 VallysLysThrLysAspThrLys 136  
Db 251 ACTGAGGCTGCTGAGGATTCTGAG 228

Search completed: October 13, 2002, 05:37:48  
Job time : 1854.09 secs

```

/note="Protein sequence is in conflict with the conceptual
translation"
/codon_start=1
/transl_table=11
/product="protein having the same IgG specificity as
protein G from group G streptococci strain G148"
/protein_id="CAA01034.1"
/db_xref="GI:4526658"
/translation="EFNKVCSDYYNNKLNNAKTVEGVKDLQAOVVESAKKARISAT
DGLSDFLSKQTPAEDTVKSTIELAEAKVLNRELDKYGSDYHKNLNNAKTVEGVKSDY
QAOVVESAKKARISATDGLSDFLSKQTPAEDTVKSTIELAEAKVLNRELDKYGSDY
YKLNINNAKTVEGVKALIDELIALPKTDYTKLLNGKTLKGTTTEANDATAKEKVF
KQVANDNGVDEMTYDDATKFTVTEKPEVIDASELTPATVTKYKLVINGKTKLGETTF
EYVADTAKEVKQYANDNGVDEMTYDDATKFTVTEKPEVIDASELTPATVTKYKLV
INGKTKLGETTFKRAVDAEAKFKQYANDNGVDMYDDATKFTVTEKPEVIDASELTP
APTEKPEKPIASIPVLPVTPATPIAKDDAKKDDTKKEDAKKPEAKKEDAKKAEPLPTTG
EGSNPFPTAAALAVMAGAGALAVASRKED"
BASE COUNT      545 a   280 c   297 g   347 t
ORIGIN

Alignment Scores:
Pred. No.:          2.68          Length:      1469
Score:          112.00          Matches:      44
Percent Similarity: 39.44%      Conservative: 12
Best Local Similarity: 30.99%    Mismatches:   62
Query Match:      14.41%        Indels:      24
DB:                6           Gaps:         5

US-09-847-539A-6 (1-159) x AL2446 (1-1469)
Qy  41  LysLysGlnAlaIleGluAspLysGluAlaThrAlaIleGlu-----Ala 56
    |||  ::|  |||:::  |||  |||  |||:::
Db  943  AAACAGAGGATCGATCGCTGCTGAATTAACACACGCGTGACCACTTACAAACTTGT 1002
Qy  57  AlaSerSerAspAlaLeu-----GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSer 74
    ::|  |||  |||  |||  |||  |||:::
Db  1003  ATTAATGGTAAACATTTGAAGGCGCAACAACTACTAAAGCAGTAGACGAGAACTGCA 1062
Qy  75  GluGluAlaValValLysAlaAspAsnAlaAlaSerAspAlaLeuGluAlaLeuAla 94
    |||:::||||  |||:::|:::  |||  ::|  ::|  ::|  ::|  ::|  ::|  ::|  ::|
Db  1063  GAAAAGCGCTCAACAATACGCTAACGACAAACGGTGTGATGGTGTGGACTTATGAT 1122
Qy  95  AspGlnThrAspAlaLeuGlnSerGluGluAlaGluValGlnSerAspAsnAlaAla 114
    |||  |||  |||  |||  |||  |||  ::|  ::|  ::|  ::|  ::|  ::|  ::|
Db  1123  GATCGACTAAGACCTTTACGCTAACTGAAATG---GTTACAGAGGTTCTCTGGTGATGCA 1179
Qy  115  SerAspAlaTrpGluLys-----AlaAlaThr 123
    |||:::|::|  |||:::|::|  |||:::|::|  |||:::|::|  |||:::|::|  |||:::|::|
Db  1180  CCAACTGAACAGAAAAACCAAGACGAAGTATCCCTCTTGTTCGGTTAACTCTCGCAACT 1239
Qy  124  ProIleAlaLeuAspValLysLysThrLysAspThr-----LysPro 137
    |||:::|::|  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db  1240  CCATTCCTAAGATGACGCTAAGAAGACGATACTAAGAAGAAGATGCTAAAAACCA 1299
Qy  138  ValValLysLysGluGluArgGlnAsnValAsnThrLeuProThrThrGlyGluGluSer 157
    |||:::|::|  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db  1300  GAAGCTAAGAAGAAGACGCTAAGAAAAGCTCAAACTCTTCTTCACTGGTGAAGCAGC 1359
Qy  158  AsnPro 159
    |||:::|::|  |||:::|::|  |||:::|::|  |||:::|::|  |||:::|::|  |||:::|::|
Db  1360  AACCCA 1365

RESULT 40
AB025967/c
LOCUS      AB025967
DEFINITION  Oryzias latipes mRNA for chorionogen Hminor, complete cds.
ACCESSION  AB025967
VERSION    AB025967.1 GI:4589849
KEYWORDS   chorionogen Hminor.
SOURCE     Oryzias latipes liver cDNA to mRNA.
ORGANISM  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;

```

Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;  
Belontiiformes; Adrianichthyidae; Oryziinae; Oryzias.

REFERENCE  
AUTHORS Sugiyama,H., Yasumasu,S., Murata,K., Tuchi,I. and Yamaami.K.  
TITLE The third egg envelope subunit in fish: cDNA cloning and analysis,  
Dev. Growth Differ. 40 (1), 35-45 (1998)  
JOURNAL 98223398

REFERENCE  
AUTHORS Sugiyama,H. and Tuchi,I.  
TITLE Direct Submission  
JOURNAL Submitted (08-APR-1999) Hitoshi Sugiyama, St. Marianna University  
School of Medicine, Department of Physiology; 2-16-1 Sugao,  
Miyamae-ku, Kawasaki, Kanagawa 216-8511, Japan  
(E-mail:sugiyamame@marianna-u.ac.jp, Tel:81-44-977-8111(ex.3405),  
Fax:81-44-977-0172)

FEATURES  
source Location/Qualifiers  
1..2025  
    /organism="Oryzias latipes"  
    /db\_xref="taxon:8090"  
    /tissue\_type="liver"  
26..1930  
    /codon\_start=1  
    /product="chorioenin Hminor"  
    /protein\_id="BAA76901.1"  
    /db\_xref="GI:459850"  
    /translation="MAMTMRLLHFCLLAVTLHNLADGDKVGQPPHYPOYPKQPONPO  
QQPQHQQPQQPQYPSKPQQPQYPNQPPQYPSKPOOHQPHQNQPOYPSRPQNPQ  
QYPPNQNPQYPSKPQQPQPHQNPQYPSKNQPPQYPSKPPQYPSKHQNDQNPQ  
POYSPKNQPPQYPSKPQQPQYPSKSNQPTPSTFHSCEVDVRYKIPGCS  
SNTASEDAINCQDHTCYGYVTLOCTKDGFILIAKDATLPHIDLESVSFLG  
SGAECHPGASSAFALQEPVPTSCGTIMKDEPCVLAYENHWSSEFYEAIGRGATRD  
SHPELSVCQRICYGTSVLETVDLVPDPGPAAGILLVELRLHGQCQVKCYCED  
VAFTSYINDADYPTVKLRDNVTVEMRLERTDNLVLTGRNGQTSSFSNVPMOD  
PLTHGPCNDDRYLVTLVPDPSGLPFTHYRRFFKMFYVGNQPTSDFSDSKKAP  
SDAPNELPLKEHVYIHCDSVYCQPTLSNNCEPLRKARKADVSGSTRKAYREETVVVYS  
DEIVFTAASP"

BASE COUNT     555 a   557 c   392 g   521 t

ORIGIN

Alignment Scores:  
Pred. No.:               3.73               Length:               2025  
Score:                  112.00              Matches:             37  
Percent Similarity:     48.44%           Conservative:          25  
Best Local Similarity:  28.91%           Mismatch:              42  
Query Match:            14.41%           Indels:                24  
DB:                      5                Gaps:                  5

US-09-847-539A-6 (1-159) x AB025967 (1-2025)

Qy   30 LeuArgAsnGluGluArgAlaIleAspGluLeuLysLysGlnAlaIleGluAspLysGlu 49  
      |||||     |||     |||     |||     |||     |||     |||     |||     |||  
Db   602 TTGAAGGATACTGAGGCTGCTGAGGACTCTGAGGCTGCTGAGGCTTTTGAGGATCGAG 543  
      |||||     |||||     |||||     |||||     |||||     |||||     |||||     |||||  
Qy   50 Ala-----ThrThrAlaIleGluAlaAlaSerAsp 60  
      |||     |||     |||     |||     |||     |||     |||     |||     |||  
Db   542 GCTGCTCAGGATTC TGAGGTTTGAAGGACTACTGAGGCTGCTGAGGATTC TGAGGCTGAT 483  
      |||     |||     |||     |||     |||     |||     |||     |||     |||  
Qy   61 ---AlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaVal 79  
      |||     |||||     |||     |||     |||     |||     |||     |||     |||  
Db   482 GAGCGCTGCTGAGGCTTTGAAGGAT--ACTGAGGCT----GCTGAGGATTC TGAGGTT 432  
      |||     |||||     |||     |||     |||     |||     |||     |||     |||  
Qy   80 VallysAlaAspAsnAlaAlaSerAspAla----- 89  
      |||     |||||     |||||     |||||     |||     |||     |||     |||     |||  
Db   431 TTGAAGGGTACTGAGGCTGCTGAGGATTC TGAGGATTC TGAGGCTGATGAGGCTTTTGAAG 372  
      |||     |||     |||     |||     |||     |||     |||     |||     |||  
Qy   90 --LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaGluVal 108  
      |||||     |||     |||     |||     |||     |||     |||     |||     |||  
Db   371 GATAC T GAGGCT GTT GAGGAT CTCTGAGGAT ACTGAGGCTGCTGAGGATCTGAGGTTTTG 312  
      |||     |||     |||     |||     |||     |||     |||     |||     |||  
Qy   109 GlnSerAspAsnAlaAlaSerAspAlaAlaTrpGluIlyAlaAlaThrProIleAlaLeuAsp 128  
      |||     |||     |||     |||     |||     |||     |||     |||     |||

```
QY 29 AlaLeuArgAsnGluGluArgAlaIleAspGluLeuLysLysGlnAlaIleGluAspLys 48
Db 969 GATGCAGCTAACCTGAAAAA-----GAAAGCTCCAAATCACAAAGCA----- 1010
QY 49 GluAlaThrThrAlaIleGluAlaAlaSerSerAspAlaLeuGluAlaLeuAlaAspGln 68
Db 1011 -----GCAGCCCTAGAAAAACAACACTAGAAAGCAACTAAAAAAGAG 1049
QY 69 ThrAspAlaLeuGlnSerGluGluAlaAlaValValLysAlaAspAsnAlaAlaSerAsp 88
Db 1050 TTAGCTGATTACAGGCTTAATAGTCAGCAACCAACCAAGAAAAAGAAAGTTAGAAAGCT 1109
QY 89 AlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuSerGlnGluAlaGluValVal 108
Db 1110 GAAGCAAAAGCTCTTAAGAGCAATTA---ACAAACAAAGCTGAAGAGCTGTGAAGCTA 1166
QY 109 GlnSerAspAsnAlaAlaSerAspAlaTrpGluLysAlaAlaThrProIleAlaLeuAsp 128
Db 1167 AAAGCAGATAAAGCTTCAGCA-----GCTCAAAAACCA-----GAT 1202
QY 129 ValLysLysThrLysAspThrLysProValValLysLysGluGluArgGlnAsnValAsn 148
Db 1203 ACCAAACCTGGCAATAAAGAGGTTCACAAAGACCGTCACAAACAAAGACAAACACTAAT 1262
QY 149 Thr-----LeuProThrThrGlyGluGlu---SerAsn 158
Db 1263 ACAGCTCTATGCGCAACAAAGAGACATATACCGTCAACAGCGCAAGAAACAAACCAAC 1322
QY 159 Pro 159
Db 1323 CCA 1325

RESULT 38
SG148IGG 1469 bp DNA linear BCT 12-JUL-1995
LOCUS Streptococcus G 148 gene fragment for IgG-binding protein G.
DEFINITION X04015
ACCESSION X04015
VERSION X04015.1 GI:47071
KEYWORDS gamma-immunoglobulin binding protein G.
SOURCE Streptococcus sp.
ORGANISM Streptococcus sp.
REFERENCE 1 (bases 1 to 1469)
AUTHORS Guss, B., Eliasson, M., Olsson, M., Uhlen, M., Frej, A.K., Jornvall, H.,
Flock, J.I. and Lindberg, M.
TITLE Structure of the IgG-binding regions of streptococcal protein G
JOURNAL EMBO J. 5 (7), 1567-1575 (1986)
MEDLINE 86300657
COMMENT Repetitive structures of 24 amino acids, A1, A2 and A3, are found
interrupted by two unrelated repeats of 51 amino acids, B1 and B2.
Further downstream, following a 12-residue unique 'spacer'-region'
(5), there is a stretch of 55 amino acids repeated three times, C1,
C2 and C3. The latter regions are interrupted by two 15-residue
regions, called D1 and D2. Directly following these repeats there
is a region W, which consists of a repetitive, extremely
hydrophilic, structure containing several proline residues. This
region resembles a similar structure in the streptococcal M protein
and in protein A, which is thought to mediate the binding of the
protein to the cell wall.
Location/Qualifiers
1..1469
/organism="Streptococcus sp."
/strain="G 148"
/db_xref="taxon:1306"
<1..1443
/codon_start=1
/transl_table=11
/product="protein G"
/protein_id="CAA27638.1"
/db_xref="GI:47072"
/db_xref="SWISS-PROT:P19909"
/translation="EFNKYGVSDYKLNINNAKTVEGKDLQAVVESAKKARISEAT"

FEATURES
source
1..1469
/organism="Streptococcus sp."
/strain="G 148"
/db_xref="taxon:1306"
<1..1443
/codon_start=1
/transl_table=11
/product="protein G"
/protein_id="CAA27638.1"
/db_xref="GI:47072"
/db_xref="SWISS-PROT:P19909"
/translation="EFNKYGVSDYKLNINNAKTVEGKDLQAVVESAKKARISEAT"

CDS
1..1443
METHOD AND MEANS FOR PRODUCING A PROTEIN HAVING THE SAME IGG
SPECIFICITY AS PROTEIN G
Patent: WO 8705631-A 4 24-SEP-1987;
Location/Qualifiers
1..1469
/organism="synthetic construct"
/db_xref="taxon:32630"
1..1443
CDS

BASE COUNT 546 a 279 c 297 g 347 t
ORIGIN
Alignment Scores: 2.68 Length: 1469
Pred. No.: 112.00 Matches: 44
Score: 39.44% Conservative: 12
Percent Similarity: 30.99% Mismatches: 62
Best Local Similarity: 14.41% Indels: 24
Query Match: 1 Gaps: 5
DB: 1

US-09-847-539A-6 (1-159) x SG148IGG (1-1469)
QY 41 LysLysGlnAlaIleGluAspLysGluAlaThrThrAlaIleGlu-----Ala 56
Db 943 AAACCAGAGTGATCGATCGCTGTAATTAAACACCGCGTACAACTTACAACTTGT 1002
QY 57 AlaSerSerAspAlaLeu-----GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSer 74
Db 1003 ATTAATCGTAAACATTGAAAGCGCAACAACTACTAAAGCAGTAGACGCAAGAACTGCA 1062
QY 75 GluGluAlaAlaValValLysAlaAspAsnAlaAlaSerAspAlaLeuGluAlaLeuAla 94
Db 1063 GAAAGAGCTTCAACCAACTACGCTAACGACACACCGGTGTGATGCTGTTGGACTATGAT 1122
QY 95 AspGlnThrAspAlaLeuGlnSerGluGluAlaGluValValGlnSerAspAsnAlaAla 114
Db 1123 GATCGGACTAAGACCTTTACGGTAACTGAAATG---GTTACAGAGGTTCCGTGGTGATGCA 1179
QY 115 SerAspAlaTrpGluLys-----AlaAlaThr 123
Db 1180 CCAACTGAACCAAGAAACCAAGCAAGATATCCCTCTGTTCCGTTAACTCTCGCAACT 1239
QY 124 ProfileAlaLeuAspValLysLysThrLysAspThr-----LysPro 137
Db 1240 CCAATTCCTAAAGATGACGCTAACAGACAGCATACTAAGAAGAAGATGCTTAAACCA 1299
QY 138 ValValLysLysGluGluArgGlnAsnValAsnThrLeuProThrThrGlyGluGluSer 157
Db 1300 GAAGCTAAGAAAGAAGACGCTAAGAAAGCTGAAAGCTCTCTCTACAACTGCTGAAGGAAGC 1359
QY 158 AsnPro 159
Db 1360 AACCCA 1365

RESULT 39
LOCUS A12446
DEFINITION 1.5 kb EcoRI/HindIII restriction fragment of pSPG1.
ACCESSION A12446
VERSION A12446.1 GI:491352
KEYWORDS synthetic construct.
SOURCE synthetic construct.
ORGANISM synthetic construct.
REFERENCE 1 (bases 1 to 1469)
AUTHORS artificial sequence.
TITLE 1 (bases 1 to 1469)
JOURNAL METHOD AND MEANS FOR PRODUCING A PROTEIN HAVING THE SAME IGG
FEATURES source
1..1469
/organism="synthetic construct"
/db_xref="taxon:32630"
1..1443
CDS
```

Alignment Scores:  
Pred. No.: 256 Length: 146491  
Score: 113.00 Matches: 37  
Percent Similarity: 47.88% Conservative: 42  
Best Local Similarity: 22.42% Mismatches: 60  
Query Match: 14.54% Indels: 26  
DB: 2 Gaps: 6

US-09-847-539A-6 (1-159) x AC019950 (1-146491)

Qy 11 IleProAsnGlycylThrLeuThrAsnLeuLeuGlyAsnAlaProGluLysLeuAlaLeu 30  
:::|||||  
Db 104506 GTCCCCAAT-----ATAAATAGTCAGCCCGAGAAATGGATGG 104468  
Qy 31 ArgAsnGlu-----GluArgAlaIleAspGluLeuLysGlnAla 44  
:::|||||  
Db 104467 GACTCAGAAGCGCGACGAGAGGCTTCGAAGCGCGAGGTCAGATCAAGAAGAGGCT 104408  
Qy 45 IleGluAspLysGluAlaThrThrAlaIleGluAlaAlaSer-----SerAsp 60  
:::|||||  
Db 104407 GAGCTTGAAACAGTCAAGAGGAATTTATCAAGAGGATTCACCAATCCGCATTCGGAT 104348  
Qy 61 AlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVal 80  
:::|||||  
Db 104347 TCGGTGGCGGAATTAAGACAGCGCGTAACCTGCTCGAGCGGAGGATGATGTCATCTA 104288  
Qy 81 LysAlaAspAsnAlaAlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeu 100  
:::|||||  
Db 104287 GAGGCGGATAAATACGCAAGAGGCTGCTAGATGATTAATGTCGCGAGCGGAG---CTA 104231  
Qy 101 GlnSerGluGluAlaGluValValGlnSer---AspAsnAlaAlaSerAspAlaTrpGlu 119  
:::|||||  
Db 104230 CCTGATCAGGAGGAGGATCTTCAATCTGAAGAGATGCGCAGACAGGCGGTAGAT 104171  
Qy 120 LysAlaAlaThr-----ProIleAlaLeuAspValLysLysThr 132  
:::|||||  
Db 104170 AGGTCAGTACCGACGACGAGTACCTGTGCGCCACCTACCGAGTATCTACCGAACAATG 104111  
Qy 133 LysAspThrLysProValLysLysGluGluArgGlnAsnValAsnThrLeuProThr 152  
:::|||||  
Db 104110 GAGATTGATGAACGACGAGCGGAGGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 104051  
Qy 153 ThrGlyGluGluSer 157  
:::|||||  
Db 104050 GAGGAGAGGAGGAGG 104036

RESULT 37  
SPU52008 4469 bp DNA linear BCT 27-AUG-1996  
LOCUS Streptococcus pyogenes emm gene cluster, mrp50, emm50 and enn50  
DEFINITION genes, complete cds.  
ACCESSION U52008  
VERSION U52008.1 GI:1354360  
KEYWORDS  
SOURCE Streptococcus pyogenes strain-B514.  
ORGANISM Streptococcus pyogenes  
Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
Streptococcus.

REFERENCE 1 (bases 1 to 4469)  
AUTHORS Yung,D.L. and Hollingshead,S.K.  
TITLE DNA sequencing and gene expression of the emm gene cluster in an  
M50 group A streptococcus strain virulent for mice  
JOURNAL Infect. Immun. 64 (6), 2193-2200 (1996)  
MEDLINE 96239037  
REFERENCE 2 (bases 1 to 4469)  
AUTHORS Hollingshead,S.K. and Yung,D.L.  
TITLE Direct Submission  
JOURNAL Submitted (21-MAR-1996) Susan K. Hollingshead, Microbiology, Univ.  
of Alabama at Birmingham, 845 19th Street South BBRB 406/6,  
Birmingham, AL 35294, USA  
FEATURES  
source Location/Qualifiers  
1..4469  
/organism="Streptococcus pyogenes"

/strain="B514"  
/db\_xref="taxon:1314"  
123..1400  
/gene="mrp50"  
123..1400  
/gene="mrp50"  
/note="IgG binding protein; Allele: 50; M-related protein"  
/codon\_start=1  
/transl\_table=11  
/product="Mrp50"  
/protein\_id="AAB06612.1"  
/db\_xref="GI:1354361"  
/translation="MSKRPNKHYSLRKLKGTASVAVALTVLGTLTNTTDDVKAETV  
GRFDEQVRKAREKAIEDVFGYTGARSVYOSGNLPNRLTPTTKSLKMQQMYKETLQK  
KEEDTLTKALTDITTEKRIESENAYKELGOLKAAAEAEAKDALNNKKQISDLT  
NNAQLKEATIEGYVOTIQNASREIAAKQOELAAAKSQAELNAETALKQODASKTEE  
TAKQSEATLENLIGLSAKHELTDIAQKLDATAPKAKLESQVTTLENLIGLSAKREL  
DLOAKLDARNAEKEKLOSOAALEKQLEATKKELADLOAKLAATNOEKEKLEAEAKAL  
KEOLYQAEELAKADKASGAKPDTAPGNKEVTPRPSOTRNTNTAPMAQTKRKQLP  
STGETTPNPFPTAAALTVIASAGVLAKRKEEN"  
1625..2866  
/gene="emm50"  
1625..2866  
/gene="emm50"  
/note="IgG binding protein; Allele: 50; M protein of  
serotype 50"  
/codon\_start=1  
/transl\_table=11  
/product="Emm50"  
/protein\_id="AAB06613.1"  
/db\_xref="GI:1354362"  
/translation="MARKDTNLQYSLRKLKGTASVAVAVAVLGVAGFANQTTVKAEEA  
GASRTITSENTISKLYDENSLIEERADLLGLEKEDKLESVERHYLRTLDQYKEH  
QKEKQOQOEEEROKYOELEROSOREVEKRYOEOLOQOOLEKEKOISEARSKSL  
RDLEASRAAKKELEAHQKLEEKQISASRQGLSRDLSEAREAKKYEADLAALTAE  
HQKLEEKQISEASRQGLSRDLSEAREAKKYEADLAALTAEHQKLEEKQISASRQ  
GLSRDLSEAREAKKYEADLAALTAEHQKLEEKQISASRQGLSRDLSEAREAKK  
AKALKQELAKQAEELAKLKGQNTPNNAKVAPOANRSRSAMTQOKRTLPTGETANPFT  
BEAATVMVSAGMLAKRKEEN"  
3077..4204  
/gene="enn50"  
3077..4204  
/gene="enn50"  
/note="IgA binding protein; Allele: 50; M-related protein"  
/codon\_start=1  
/transl\_table=11  
/product="Enn50"  
/protein\_id="AAB06614.1"  
/db\_xref="GI:1354363"  
/translation="MARQTKTKYSLRKLKGTASVAVAVALTVLGAGFANQTEVRADQA  
ASGTSVNNGASTSVESFKYDALRGENDLRNVNAKYLEKINAEKKNKLEATNNKELN  
ENYKLOPDIDALEKEKEDSKTTLAKTKENEISPAKGLSRDLSEAREAKKYEADLA  
HOKLEAKNKLTEANQISEASRQGLSRDLSEAREAKKYEADLAHOKLEAKHOKLEAK  
KLAGEHVSSTSRKGLSRDLSEAREAKKYSSELTOAKAQLSALSEKSKLEKAEEL  
QAKLDAQCKALKQELAKLKEAKGAGGTPATKANKRSRSGRAAQTATRPSPQ  
NKGMRSQLPSTGEAANPFTTAAATVMVSAGMLTKRKEEN"  
BASE COUNT 1841 a 860 c 870 g 898 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 7.72 Length: 4469  
Score: 112.50 Matches: 49  
Percent Similarity: 42.24% Conservative: 19  
Best Local Similarity: 30.43% Mismatches: 54  
Query Match: 14.48% Indels: 39  
DB: 1 Gaps: 8  
US-09-847-539A-6 (1-159) x SPU52008 (1-4469)  
Qy 16 ThrLeuThrAsnLeuLeuGlyAsnAlaProGlu-----LysLeu 28  
|||||  
Db 909 ACCTTAGAAACCTCCTAGGTTACAGCTAAGCGTGATGACTGATCTGCAAGCTAAGCTA 968  
|||||



```

Qy 59 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAla 78
      ::::||||| ||||| ||||| :::: ||||| ||||| |||||
Db 206877 GCAGAGCGAGGAGCGAGGAGCGAGGAGCGAGGAGCGAGGAGCGAGGAGCGAGGAG 206827
      ::::||||| ||||| ||||| :::: ||||| ||||| |||||
Qy 79 ValValAlaAspAsnAlaAlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAsp 98
      ::::||||| ||||| ||||| :::: ||||| ||||| |||||
Db 206826 GCAGAGGAGCGAGGAGCGAGGAGCGAGGAGCGAGGAGCGAGGAGCGAGGAGCGAGGAG 206767
      ::::||||| ||||| ||||| :::: ||||| ||||| |||||
Qy 99 AlaLeuGlnSerGluGluAlaGluValValGlnSerAspAsnAlaAlaSerAspAlaTrp 118
      ::::||||| ||||| ||||| :::: ||||| ||||| |||||
Db 206766 GCA---GAGGAGCGAGGAGCGAGGAGCGAGGAGCGAGGAGCGAGGAGCGAGGAGCGAGGAG 206710
      ::::||||| ||||| ||||| :::: ||||| ||||| |||||
Qy 119 GluLysAlaAlaThrProLeAlaLeuAspValLysThrLysAspThrLysProval 138
      ::::||||| ||||| ||||| :::: ||||| ||||| |||||
Db 206709 GGATCACTACAAATTCAGTCAGTCTTATGTCACAAACATATCATCTACAAAGAGTCTC 206650
      ::::||||| ||||| ||||| :::: ||||| ||||| |||||
Qy 139 ValLysLysGlu 142
      ::::|||||
Db 206649 TCAAGGAGCGAG 206638

RESULT 35
AC020602/c
LOCUS Homo sapiens BAC clone RP11-461M18 from 2, complete sequence.
DEFINITION AC020602
ACCESSION AC020602
VERSION AC020602.6 GI:17386508
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 167227)
AUTHORS Sulston,J.E. and Waterston,R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
REFERENCE 2 (bases 1 to 167227)
AUTHORS Waligorski,J., Cotton,M. and Doeber,A.
TITLE The sequence of Homo sapiens BAC clone RP11-461M18
JOURNAL Unpublished (2002)
REFERENCE 3 (bases 1 to 167227)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (05-JAN-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 4 (bases 1 to 167227)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (06-DEC-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 5 (bases 1 to 167227)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (07-DEC-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 6 (bases 1 to 167227)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (03-JAN-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 7 (bases 1 to 167227)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (10-JAN-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Dec 6, 2001 this sequence version replaced gi:14029092.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC

```

Web site: <http://genome.wustl.edu/gsc>  
 Contact: [sapiens@wustl.wustl.edu](mailto:sapiens@wustl.wustl.edu)  
 ----- Summary Statistics  
 Center project name: H\_NH0461M18  
 -----

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

#### MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

#### SOURCE INFORMATION:

The RP11-11 human BAC library was made from the blood of one male donor, as described by Osogawa,K., Woon,P.Y., Zhao,B., Frengen,E., Tateno,M., Catanesi,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)  
 VECTOR: pBACe3.6

#### NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-341H1, 2000 bp overlap; the clone sequenced to the right is RP11-289K3. Actual start of this clone is at base position 162181 of RP11-341H1; actual end is at base position 167227 of RP11-461M18.

Data from AC016725 and AC017031 was used to finish this clone, AC020602. There is a tandem repeat from base position 33099 to 33933.

FEATURES	source	Location/Qualifiers
	1..167227	/organism="Homo sapiens"
		/db_xref="taxon:9606"
		/chromosome="2"
		/map="2"
		/clone_lib="RP11-461M18"
		/clone_lib="RP11-11"
	250..316	/rpt_family="MIR"
repeat_region	878..923	/rpt_family="MIR"
repeat_region	1376..1470	/rpt_family="MIR"
repeat_region	2361..2391	/rpt_family="(TTTTA)n"
repeat_region	2850..3229	/rpt_family="MaLR"
repeat_region	3260..3347	/rpt_family="ERV1"
repeat_region	3488..3516	/rpt_family="(TTTTG)n"
repeat_region	3490..3796	/rpt_family="Alu"
repeat_region	5077..5106	/rpt_family="AT-rich"
repeat_region	5090..5289	/rpt_family="L1"
repeat_region	6473..6782	

# TITLE JOURNAL REFERENCE AUTHORS

High Throughput Mouse Sequencing  
2 (bases 1 to 222304)  
Montgomery,K.T., Grills,G., Li,L., Brown,W.A., Chiu,D., Decker,J.,  
Fusina,M., Halder,A., Keller,A., Perera,A., Shih,C., Thomas,E.,  
Zeneck,W., Xi,C., Juelis,P. and Kucheralapati,R.  
Direct Submission  
Submitted (21-DEC-2001) Harvard Partners Center for Genetics and  
Genomics, Harvard Medical School, 65 Landsdowne St, Cambridge, MA  
02139, USA

## COMMENT

On Jan 1, 2002 this sequence version replaced gi:17985888.  
-----Genome Center  
Center: Harvard Partners Genome Center  
Center Code: HPGC  
Web site:  
http://wchanning.bwh.harvard.edu:9088/hpcgg/jsp/hpcgg/Sequence/mous  
e.html  
Contact: gnkm@capecod.bwh.harvard.edu  
-----Summary Statistics  
Center project name: ALG  
Sequencing vector: pSMART; AF399742  
Chemistry: Dye-terminator Big Dye; 100%  
\*Consensus quality: 213980 at least Q20  
\*Consensus quality: 211926 at least Q30  
\*Consensus quality: 209034 at least Q40  
\*Estimated insert size: agarose-PP - N/A  
\*\*Estimated insert size: 221944 - sum-of-contigs  
Quality coverage: agarose-PP - N/A  
Quality coverage: 7.6 x in Q20 bases; sum-of-contigs estimation  
-----

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 19 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 44872: contig of 44872 bp in length  
\* 44873 44892: gap of unknown length  
\* 44893 67114: contig of 22222 bp in length  
\* 67115 67134: gap of unknown length  
\* 67135 86837: contig of 19703 bp in length  
\* 86838 86857: gap of unknown length  
\* 86858 103978: contig of 17121 bp in length  
\* 103979 103998: gap of unknown length  
\* 103999 121819: contig of 17821 bp in length  
\* 121820 121839: gap of unknown length  
\* 121840 133571: contig of 11732 bp in length  
\* 133572 133591: gap of unknown length  
\* 133592 149349: contig of 15758 bp in length  
\* 149350 149369: gap of unknown length  
\* 149370 157963: contig of 8594 bp in length  
\* 157964 157983: gap of unknown length  
\* 157984 167921: contig of 9938 bp in length  
\* 167922 167941: gap of unknown length  
\* 167942 178841: contig of 10900 bp in length  
\* 178842 178861: gap of unknown length  
\* 178862 192843: contig of 13982 bp in length  
\* 192844 192863: gap of unknown length  
\* 192864 200562: contig of 7699 bp in length  
\* 200563 200582: gap of unknown length  
\* 200583 208726: contig of 8144 bp in length  
\* 208727 208746: gap of unknown length  
\* 208747 212290: contig of 3544 bp in length  
\* 212291 212310: gap of unknown length  
\* 212311 215133: contig of 2823 bp in length  
\* 215134 215153: gap of unknown length  
\* 215154 218142: contig of 2989 bp in length  
\* 218143 218162: gap of unknown length  
\* 218163 219720: contig of 1558 bp in length  
\* 219721 219740: gap of unknown length  
\* 220986: contig of 1246 bp in length

\* 220987 221006: gap of unknown length  
\* 221007 222304: contig of 1298 bp in length.  
FEATURES  
Source  
1..222304  
/organism="Mus musculus"  
/strain="C57BL6/J"  
/db\_xref="taxon:10090"  
/chromosome="15"  
/clone="Rp23-274F23"  
/sex="male"  
1..44872  
/note="assembly\_name:Contig67"  
44893..67114  
/note="assembly\_name:Contig66"  
67135..86837  
/note="assembly\_name:Contig65"  
86858..103978  
/note="assembly\_name:Contig64"  
103999..121819  
/note="assembly\_name:Contig63"  
121840..133571  
/note="assembly\_name:Contig62"  
133592..149349  
/note="assembly\_name:Contig61"  
149370..157963  
/note="assembly\_name:Contig60"  
157984..167921  
/note="assembly\_name:Contig59  
clone\_end:77  
vector\_side:right"  
167942..178841  
/note="assembly\_name:Contig58"  
178862..192843  
/note="assembly\_name:Contig57"  
192864..200562  
/note="assembly\_name:Contig56"  
200583..208726  
/note="assembly\_name:Contig55"  
208747..212290  
/note="assembly\_name:Contig54"  
212311..215133  
/note="assembly\_name:Contig53"  
215154..218142  
/note="assembly\_name:Contig52  
clone\_end:SP6  
vector\_side:left"  
218163..219720  
/note="assembly\_name:Contig51"  
219741..220986  
/note="assembly\_name:Contig50"  
221007..222304  
/note="assembly\_name:Contig49"  
440 others  
BASE COUNT 56547 a 55439 c 54306 g 55572 t  
ORIGIN

Alignment Scores:  
Pred. No.: 301 Length: 222304  
Score: 114.50 Matches: 36  
Percent Similarity: 50.81% Conservative: 27  
Best Local Similarity: 29.03% Mismatches: 50  
Query Match: 14.74% Indels: 11  
DB: 2 Gaps: 3

US-09-847-539A-6 (1-159) x AC104834 (1-222304)

QY 26 GluLysLeuAlaLeuArgAsnGluGluArgAlaIleAspGluLeuLysLysGlnAlaIle 45

||| ||| ::||| |||::||| ||| |||::||| ::|  
Db 206997 GAGCACCCTAGCGGTCCGGCCAGAGAGAGGATGCACAGAGGGGAGGAGGACAGACAGTA 206998

QY 46 GluAspLysGluAla-----ThrThrAlaIleGluAlaIaSer 58

||||: |||  
Db 206937 CAAGAAGTCAGATCTGTGTCATGACTATACCTCAATACGCAGGCGAGGAGGAGGAG 206878





misc\_feature EADGDAIRKRAQAEADAIKLRGEALRONPNVMELEINKWNGOLPOVMTGANTPE  
TALK"  
complement(2427..2459)  
/gene="HCM2.0006c"  
/note="PS00013 Prokaryotic membrane lipoprotein lipid  
attachment site"  
complement(2493..2693)  
/gene="HCM2.0007c"  
complement(2493..2693)  
/gene="HCM2.0007c"  
/note="HCM2.0007c"  
highly similar to TR:CAB62372 (EMBL:AL117211), YPMW1.49AC,  
versinia pestis CO-92 hypothetical protein from plasmid  
pWT1 (66 aa), fasta scores; E(): 3.5e-25, 97.0% identity  
in 66 aa overlap. Contains hydrophobic, possible

## Alignment Scores:

Pred. No.:	141	Length:	106516
Score:	114.50	Matches:	40
Percent Similarity:	46.51%	Conservative:	20
Best Local Similarity:	31.01%	Mismatches:	44
Query Match:	14.74%	Indels:	14 25
DB:	1	Gaps:	5

US-09-847-539A-6 (1-159) x STVPPHCM2 (1-106516)

QY 37 IleAspGluLeuLysLysGlnAlaIleGluAspLysGluAlaThrThrAlaIleGluAla 56  
:::||||| ::::: ::::: ||| |||||  
Db 39784 GTCGATACCATTCAGTCGGAGTCTCAAAAGGCTCTGGCTGTCGCGCCGACGCGAAGCA 39725  
||| ::::: |||||  
QY 57 AlaSerSerAspAlaLeu-----GluAlaLeu 65  
||| ::::: |||||  
Db 39724 CGCGCTGCAGAGCGGTAACTCTAAACAGACAGCTCGCGGCGGAGCGAAGCTAATGCTTAAG 39665  
||| ::::: |||||  
QY 66 AlaAspGlnThrAspAlaLeuGlnSerGluGluAlaValValLysAlaAspAsnAla 85  
||| ::::: |||||  
Db 39664 GCAAGTGAAACAGCTGCAGCTGCCTCTCAGCAGCGAGCGCTACCAGTGAAGTAAGGCC 39605  
||| ::::: |||||  
QY 86 AlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAla 105  
::: ||| ||| ||||| |||||  
Db 39604 CGGCGAAGCAAGAGCGCTCGCGAGCATCACAAACAGCTCATTAACAAAGTACGAGGTT 39545  
||| ::::: |||||  
QY 106 GluValValGlnSerAspAsnAlaAlaSerAspAlaThrGluLysAlaAlaThrProIle 125  
||| ::::: |||||  
Db 39544 GCAGCGCCCATCGACACTCGCGAAGAGTGAAGCA---GAAAGGCG----- 39500  
||| |||||  
QY 126 AlaLeuAspValLysLys-----ThrLysAspThrLysProValValLysLysGlu 142  
||| |||||  
Db 39499 -----AAGAAATTTGCCGATCTCTTAGATGTAACAACAGCTGCTGCATAAAGAT 39452  
||| |||||  
QY 143 GluArgGlnAsnValAsnThrLeuPro 151  
||| |||||  
Db 39451 -----CAGAACCTTGCTGACCTGCCA 39431  
||| |||||  
RESULT 33  
AC023354  
LOCUS  
DEFINITION Mus musculus chromosome 15 clone RP23-384L8 strain C57BL6/J,  
WORKING DRAFT SEQUENCE, 29 unordered pieces.  
ACCESSION AC023354  
VERSION 1  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 210791)  
AUTHORS Montgomery,K.T., Grills,G., Han,J., Lee,E., Long,J., Pomerantz,R.,  
Toshikhes,I.P., Shim,C., Decker,J., Thomas,E., Perera,A.,  
Gordon,M., Goltz,J.S. and Kucherlapati,R.  
TITLE High Throughput Mouse Sequencing  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 210791)

## AUTHORS

Montgomery,K.T., Grills,G., Han,J., Lee,E., Long,J., Pomerantz,R.,  
Toshikhes,I.P., Shim,C., Decker,J., Thomas,E., Perera,A.,  
Gordon,M., Goltz,J.S. and Kucherlapati,R.

TITLE  
JOURNAL

Direct Submission  
Submitted (14-FEB-2000) Department of Molecular Genetics, Albert  
Einstein College of Medicine Genome Center, 1300 Morris Park Ave.,  
Bronx, NY 10461, USA

## COMMENT

On Jan 9, 2002 this sequence version replaced gi:11094415.  
-----Genome Center  
Center: Harvard Partners Genome Center  
Center Code: HPGC

Web site:  
http://wchanning.bwh.harvard.edu:9088/hpcgg/jsp/hpcgg/Sequence/mous  
e.html

Contact: gkntm@ceped.bwh.harvard.edu  
-----Summary Statistics

Center project name: ABA  
Sequencing vector: pUC18; L08752  
Chemistry: Dye-terminator Big Dye; 100%

\*Consensus quality: 198418 at least Q20

\*Consensus quality: 195838 at least Q30

\*Consensus quality: 191563 at least Q40

Estimated insert size: agarose-PP - N/A

\*\*Estimated insert size: 210231 - sum-of-contigs

Quality coverage: agarose-PP - N/A

Quality coverage: 4.8 x in Q20 bases; sum-of-contigs estimation

-----

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 29 contigs. The true order of the pieces

\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will

\* be preserved.

\* 1 21856: contig of 21856 bp in length

\* 21857 21876: gap of unknown length

\* 21877 48688: contig of 26812 bp in length

\* 48689 48708: gap of unknown length

\* 48709 65965: contig of 17257 bp in length

\* 65966 65985: gap of unknown length

\* 65986 80051: contig of 14066 bp in length

\* 80052 80071: gap of unknown length

\* 80072 94631: contig of 14560 bp in length

\* 94632 94651: gap of unknown length

\* 94652 108022: contig of 13371 bp in length

\* 108023 108042: gap of unknown length

\* 108043 119443: contig of 11400 bp in length

\* 119443 119462: gap of unknown length

\* 119463 132361: contig of 12899 bp in length

\* 132362 132381: gap of unknown length

\* 132382 141581: contig of 9200 bp in length

\* 141582 141601: gap of unknown length

\* 141602 151955: contig of 10354 bp in length

\* 151956 151975: gap of unknown length

\* 151976 159593: contig of 7618 bp in length

\* 159594 159613: gap of unknown length

\* 159614 167366: contig of 7753 bp in length

\* 167367 167386: gap of unknown length

\* 167387 173218: contig of 5832 bp in length

\* 173219 173238: gap of unknown length

\* 173239 179907: contig of 6669 bp in length

\* 179908 179927: gap of unknown length

\* 179928 185771: contig of 5844 bp in length

\* 185772 185791: gap of unknown length

\* 185792 189317: contig of 3526 bp in length

\* 189318 189337: gap of unknown length

\* 189338 193220: contig of 3883 bp in length

\* 193221 193240: gap of unknown length

\* 193241 196107: contig of 2867 bp in length

\* 196108 196127: gap of unknown length

\* 196128 199286: contig of 3159 bp in length

\* 199287 199306: gap of unknown length

```

/misc_feature
/db_xref="taxon:90370"
complement(1..29)
/gene="HCM2.0001c"
CDS
complement(1..29)
/partial
/gene="HCM2.0001c"
/notes="HCM2.0001c, hypothetical protein, len: 68 aa; highly similar to TR:CAB55227 (EMBL:AL117211), YPWT1.45C, Yersinia pestis CO-92 hypothetical protein from plasmid pMT1 (71 aa), fasta scores: E(): 5.9e-25, 93.3% identity in 60 aa overlap and TR:O68760 (EMBL:AF074611), Y1119, Yersinia pestis KIM5 hypothetical protein from plasmid pMT1 (71 aa), fasta scores: E(): 5.9e-25, 93.3% identity in 60 aa overlap. Spans the end of the sequence"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAD09868.1"
/db_xref="GI:16505982"
/translacion="MAVTLAGLE"
1..232
/notes="95.7% identical to Yersinia pestis CO-92 plasmid pMT1 (EMBL:AL117211) bases 48304..48535"
complement(29..334)
/gene="HCM2.0002c"
CDS
complement(29..334)
/gene="HCM2.0002c"
/notes="HCM2.0002c, possible periplasmic protein, len: 101 aa; similar to TR:CAB55228 (EMBL:AL117211), YPWT1.46C, Yersinia pestis hypothetical protein from plasmid pMT1 (111 aa), fasta scores: E(): 2.8e-26, 95.5% identity in 67 aa overlap. The N-terminal 37 aa are highly similar to YPWT1.46AC (EMBL:AL117211) Yersinia pestis hypothetical protein from plasmid pMT1 (59 aa) (96.8% identity in 38 aa overlap) which lies immediately upstream of YPWT1.46C in pMT1. Contains a possible N-terminal signal sequence"
/codon_start=1
/transl_table=11
/product="putative periplasmic protein"
/protein_id="CAD09869.1"
/db_xref="GI:16505983"
/translacion="MKTVDLTLLIVSIAFVLCIFGTGATRKALAPVNGAMVVDSEGQ FYVVRDRLVSSPLTKRNGKLSGVGDREYFNKSGDVYGVGHAKNASYLPDDCKEVG"
complement(36..43)
/gene="HCM2.0002c"
/notes="possible RBS"
224..6722
/notes="96.5% identical to Yersinia pestis CO-92 plasmid pMT1 (EMBL:AL117211) bases 48733..55208"
complement(375..650)
/gene="HCM2.0003c"
CDS
complement(375..650)
/gene="HCM2.0003c"
/notes="HCM2.0003c, hypothetical protein, len: 91 aa; highly similar to TR:CAB62371 (EMBL:AL117211), YPWT1.46BC, Yersinia pestis CO-92 hypothetical protein from plasmid pMT1 (91 aa), fasta scores: E(): 0.94.5% identity in 91 aa overlap and TR:O92GX9 (EMBL:AF074611), Y1117, Yersinia pestis KIM5 hypothetical protein from plasmid pMT1 (91 aa), fasta scores: E(): 0.94.5% identity in 91 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAD09870.1"
/db_xref="GI:16505984"
/translacion="MKILVRISASTDYVDYVPLFMVKCDGLNDEFIQAAIERNLVEYTG MDADSVYDDGVGVHNGSCWYVDDTMTVSDDEAAHLERILIGISTFE"
complement(658..663)
/notes="possible RBS"
complement(719..1129)
/gene="HCM2.0004c"
CDS
complement(719..1129)

```

---

```

/gene="HCM2.0004c"
/notes="HCM2.0004c, hypothetical protein, len: 136 aa; highly similar to TR:CAB55229 (EMBL:AL117211), YPWT1.47C, Yersinia pestis CO-92 hypothetical protein from plasmid pMT1 (136 aa), fasta scores: E(): 0.97.8% identity in 136 aa overlap and TR:O68762 (EMBL:AF074611), Y1116, Yersinia pestis KIM5 hypothetical protein from plasmid pMT1 (136 aa), fasta scores: E(): 0.97.8% identity in 136 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAD09871.1"
/db_xref="GI:16505985"
/translacion="MKKLYDAANAALDVVDTETIAOQGFPEPEMATQLEATAENNAPEP SEEDADQRFIRMAEVEGPTPTAEQAMLLKYFKEAGENLPVDDTPHFWHAARWKFV IYTRGSGKDMVWHLMDKAVDRYLEKFFPPA"
complement(1113..1484)
/gene="HCM2.0005c"
CDS
complement(1113..1484)
/gene="HCM2.0005c"
/notes="HCM2.0005c, hypothetical protein, len: 123 aa; similar to TR:Q38665 (EMBL:X78401), ninX, Bacteriophage P22 NinX protein (112 aa), fasta scores: E(): 8.9e-15, 45.6% identity in 114 aa overlap. Highly similar to TR:CAB55230 (EMBL:AL117211), YPWT1.48C, Yersinia pestis CO-92 hypothetical protein from plasmid pMT1 (123 aa), fasta scores: E(): 0.87.8% identity in 123 aa overlap and TR:O92GY0 (EMBL:AF074611), Y1115, Yersinia pestis KIM5 hypothetical protein from plasmid pMT1 (123 aa), fasta scores: E(): 0.87.8% identity in 123 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAD09872.1"
/db_xref="GI:16505986"
/translacion="MKNYSEMTDFEINCLVAETGHRPLISQYGVKWSQBGDYTAVVA IGPNGAGTFDWCNDEPDANDIYRHRIGVIPARQPGEMRAHRKVDSSSTPQHLIONPN PFRAAMTVFLMQEKKHEETV"
complement(1137..1141)
/gene="HCM2.0005c"
/notes="possible RBS"
complement(1492..1497)
/notes="possible RBS"
1565..1621
/notes="repeat 1: 57 bp repeat"
complement(1647..2489)
/gene="HCM2.0006c"
CDS
complement(1647..2489)
/gene="HCM2.0006c"
/notes="HCM2.0006c, possible lipoprotein, len: 280 aa; shows weak similarity to TR:AAF10061 (EMBL:AE001907), DR0482, Deinococcus radiodurans B-cell receptor associated protein-related protein (328 aa), fasta scores: E(): 4.9e-07, 23.4% identity in 265 aa overlap. Highly similar to TR:CAB55231 (EMBL:AL117211), YPWT1.49C, Yersinia pestis CO-92 putative lipoprotein from plasmid pMT1 (276 aa), fasta scores: E(): 0.97.5% identity in 280 aa overlap and TR:O68763 (EMBL:AF074611), Y1114, Yersinia pestis KIM5 hypothetical protein from plasmid pMT1 (276 aa), fasta scores: E(): 0.97.5% identity in 280 aa overlap. Contains a possible N-terminal signal sequence and an appropriately positioned PS00013 Prokaryotic membrane lipoprotein lipid attachment site"
/codon_start=1
/transl_table=11
/product="putative lipoprotein"
/protein_id="CAD09873.1"
/db_xref="GI:16505987"
/translacion="MKKGLLAVALTACTMGLTSCDRVEPGYVGIKYNKLGEDKGICE VVGVRQWGTGLNTELYFTPTFKOMKTYDEPTFQMSDGTAGHKIGVAYLVNRDKVTT VFQYRKGVDDITDTRQKIADSLNRLASRMTDTDFIDGKASLLDNALKDQAEMS PVGIEVLSLSWVGKPDYPTDPTVIESINAKNTLQROQQEVEQRKAENMLREQAG

```

Db 94243 CAACCTGCGCCCAACTACTAACACCTTGAAGACTGCGCACTGATTCGAACAAC 94193

## RESULT 31

## STRMAG

LOCUS STRMAG 1555 bp DNA linear BCT 16-AUG-1994  
 DEFINITION Streptococcus dysgalactiae (mag) gene, complete cds.  
 ACCESSION L27798  
 VERSION L27798.1 GI:475117  
 KEYWORDS mag gene.  
 SOURCE Streptococcus dysgalactiae DNA.  
 ORGANISM Streptococcus dysgalactiae  
 Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 Streptococcus.  
 REFERENCE 1 (bases 1 to 1555)  
 AUTHORS Jonsson,H., Frykberg,L., Rantamaeki,L. and Guss,B.  
 TITLE MAG, a novel plasma protein receptor from Streptococcus  
 dysgalactiae  
 JOURNAL J. Clin. Microbiol. 32:1555-1561 (1994)  
 MEDLINE 94259307  
 FEATURES

## source

Location/Qualifiers

1..1555  
 /organism="Streptococcus dysgalactiae"

/db\_xref="taxon:1334"

## -35\_signal

230..235

## -10\_signal

249..254

## -10\_signal

264..269

## RBS

276..281

## gene

288..1529

## CDS

288..1529

/function="binding of macroglobulin, albumin and IgG"

/codon\_start=1

/transl\_table=1

/protein\_id="AAA26921.1"

/db\_xref="GI:475118"

/translation="MEKSKVKYFLRKSAFLGSLVSAFLVGTAVVNAEESTVSPVT  
 ATDVTTKEALAIINKLSEDLNLDIQEVLAKAGRDILASDSADTKALLAEVTA  
 VTRNEEKWARDVDKAIADAAFAFSELKDLQKAYEDLAKLAADTDLDLVAKIIND  
 YTKVENAKTEDVKVKEFESQNEVTRIKTERKALAKAKADAIEILKYGIGDY  
 YKLIINNKTAGVTLAKDEILASKPAVIDAPELPTALTYKLVLINGTLKGETTTKA  
 VDAETAEKAFVYANENGVDGWTYDDATKFTVEMVTEVPDGAPEPKPEASIP  
 VLPATPIAKDDAKDKDTKKDDTKKEDAKKPEAKKEAKAATLTPTTGESENPFFTA  
 AALAVMAGADALAVASKRKE"

## sig\_peptide

288..389

/gene="mag"

/note="putative"

390..1526

/gene="mag"

/note="putative"

/function="binding of macroglobulin, albumin and IgG"

/product="unnamed"

BASE COUNT 582 a 242 c 305 g 426 t

## ORIGIN

## Alignment Scores:

Pred. No.: 1.83 Length: 1555  
 Score: 114.50 Matches: 49  
 Percent Similarity: 34.22% Conservative: 15  
 Best Local Similarity: 26.20% Mismatches: 56  
 Query Match: 14.74% Indels: 6  
 DB: 1 Gaps: 6

US-09-847-539a-6 (1-159) x STRMAG (1-1555)

Qy 36 AlaIleAspGluLeuLysLysGlnAlaIleGluAsp----- 47

Db 903 GCTATTGAATTCGAGAAATACGGAATTCGCGATTACTATATAATTAATAAT 962

Qy 48 -----LysGluAlaThrThrAlaIleGlu-----AlaAlaSerSerAspAla 61

Db 963 GGTAAACTGCAGAGGCTGCTCTCTTAAAGATGAATTTTAGCTTCAAAACGACGA 1022

Qy 62 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLys 81  
 Db 1023 GTGATTGACGCACCTGAAATTACACCGCTTTTGACAACTTACAACTTGTATC----- 1076  
 Qy 82 AlaAspAlaAlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGln 101  
 Db 1077 -----AATGGTAAACATTTGAAAGCGCAACACTACTAAAGCAGTAGACGCAAACT 1130  
 Qy 102 SerGluGluAla---GluValValGlnSerAspAlaAlaSerAspAlaThrGlu--- 119  
 Db 1131 GCAGAAAAAGCCCTTCAACAACTACGCTAAGCAAAAGCGTGTGATGGTGTGGACTTAC 1190  
 Qy 119 ----- 119  
 Db 1191 GATGATGCCACTAAGACCTTTTACTGTAAGTGAATGGTTACTGAGAGTTCTGCTGTGATGCA 1250  
 Qy 120 -----LysAlaAlaThr 123  
 Db 1251 CCAACTGAACCAAAAAACCAGAACGATATCCCTCTTTGTCCTTAACTCCTCGCAACT 1310  
 Qy 124 ProIleAlaLeuAspValLysLysThrLysAspThr----- 135  
 Db 1311 CCAATTTGCTAAAGATGACGCTAAGAAAGACGATACTAAGAAAGACGATACTAAGAAAGAA 1370  
 Qy 136 -----LysProValValLysLysGluGlnAsnValAsnValAsnThrLeuProThr 152  
 Db 1371 GATGCTAAAAACCCAGAACGCTAAGAAAGAAAGACGCTAAGAAAGCTCAACTCTCTCTACA 1430  
 Qy 153 ThrGlyGluGluSerAsnPro 159  
 Db 1431 ACTGGTGAAGGAAGCAACCCA 1451  
 RESULT 32  
 STYPPHCM2/c 106516 bp DNA linear BCT 26-OCT-2001  
 LOCUS Salmonella enterica serovar Typhi (Salmonella typhi) strain CT18  
 DEFINITION plasmid pHC2.  
 ACCESSION AL513384  
 VERSION AL513384.1 GI:16505981  
 KEYWORDS Salmonella enterica subsp. enterica serovar Typhi.  
 SOURCE Salmonella enterica subsp. enterica serovar Typhi  
 ORGANISM Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 Salmonella.  
 REFERENCE 1 (bases 1 to 106516)  
 AUTHORS Parkhill,J., Dougan,G., James,K.D., Thomson,N.R., Pickard,D.,  
 Wain,J., Churcher,C., Mungall,K.L., Bentley,S.D., Holden,M.T.G.,  
 Sebaihia,M., Baker,S., Basham,D., Brooks,K., Chillingworth,T.,  
 Connor,P., Cronin,A., Davis,P., Davies,R.M., Dowd,L., White,N.,  
 Farrar,J., Feltwell,T., Hamlin,N., Haque,A., Hien,T.T., Holroyd,S.,  
 Jagels,K., Krogh,A., Larsen,T.S., Leather,S., Moule,S., O'Gaora,P.,  
 Parry,C., Quail,M., Rutherford,K., Simmonds,M., Skelton,J.,  
 Stevens,K., Whitehead,S. and Barrell,B.G.  
 Complete genome sequence of a multiple drug resistant Salmonella  
 enterica serovar Typhi CT18  
 TITLE Nature 413 (6858), 848-852 (2001)  
 JOURNAL 11677608  
 PUBMED 11677608  
 REFERENCE 2 (bases 1 to 106516)  
 AUTHORS Parkhill,J.  
 TITLE Direct Submission  
 JOURNAL Submitted (25-OCT-2001) Submitted on behalf of the Salmonella  
 sequencing team, Sanger Centre, Wellcome Trust Genome Campus,  
 Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk  
 NOTES: Details of S. typhi sequencing at the Sanger Centre are available  
 on the World Wide Web.  
 (URL, [http://www.sanger.ac.uk/projects/S\\_typhi/](http://www.sanger.ac.uk/projects/S_typhi/)).  
 FEATURES  
 source  
 1..106516  
 /organism="Salmonella enterica subsp. enterica serovar  
 Typhi"  
 /plasmid="pHC2"  
 /strain="CT18"

```
Db 13433 CGCATCGCTCACACATCTCGCCACATTTGGTTGAAGCTGTCAACGATGTTCCAGCCAGTGTT 13492
Qy 25 ---ProGluLysLeuAlaLeuArg-
|||
:::|||||
13493 GAGCCTTCAGCCGTCGCTTTTCGCGTGCCTGAAATTCGTTCTGATCTTGAAGAAG 13552
Qy 32 -----AsnGluGluArgAlaIleAspGluLeuLysLysGlnAla--- 44
:::|||||
13553 AACATCAAAAGACTGTTGACGATGAAAGCGCCGATGAACCTGAAGACGATGTCGGA 13612
Qy 45 -----IleGluAsp-----LysGluAlaThrAlaIle 54
:::|||||
13613 AATCGGTCAAGACGTCGGAAGATGTTGTTTCGAAATACCAAAATCAACCTCAACCACTC 13672
Qy 55 GluAlaAlaSerSerAspAla-----LeuGluAlaLeu 65
::| ||| |||||
13673 GACGTGGCTAAGGACGACGCTTAACAACTCAAGCTACTGTAGACCACTTACCAAGCTG 13732
Qy 66 AlaAspGlnThrAspAlaLeuGlnSerGluGluAlaValValLysAlaAspAla 85
|||::: |||||
13733 GCCGAGTCTACGACAAAGATGATCCCAAGTCGCAAGTCGCAAGGACATCAAGGACTCAAGACT 13792
Qy 86 AlaSer-----AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGln 101
::: ||| |||||
13793 AAGCTTAAGGAATTCCTTCAGCTCTTGAANAAGCCATTCACCAAGAGAGCGCTATTTCGT 13852
Qy 102 SerGluGluAlaGluValValGlnSerAspAlaAlaSerAspAlaTrpGluLysAla 121
|||::: |||||
13853 CGTGAACAGCGGAGATT-----AATGATCGTCTCAACAACTCTCGAAGGAG 13900
Qy 122 AlaThrProIleAlaLeuAspValLysThrLysAspThrLysProValValLys 141
|||::: ||| |||
13901 CTTACCAAGGTG-----GATGAGTTCAAACGAGAGATGCACCTCCCAATGTT---GAC 13951
Qy 142 GluGluArgGlnAsnValAsnThrLeuProThrThrGlyGluGluSerAsn 158
::: ||| |||||
Db 13952 CAATCGCCGCCAATACTACACCTTGAAGACTGCCACTGATTCGCAACAC 14002

RESULT 30
AC006902/c
LOCUS
DEFINITION
AC006902
Caenorhabditis elegans clone Y74A11Y, *** SEQUENCING IN PROGRESS
***, 11 unordered pieces.
AC006902
AC006902.2 GI:4309900
HTG: HTGS_PHASE1.
KEYWORDS
Caenorhabditis elegans.
SOURCE
Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditoidae; Rhabditidae; Peloderinae; Caenorhabditis.
Waterston,R.H.
The sequence of Caenorhabditis elegans clone
Unpublished
2 (bases 1 to 96468)
Waterston,R.H.
Direct Submission
Submitted (24-FEB-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Mar 1, 1999 this sequence version replaced gi:4263446.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 2040: contig of 2040 bp in length
* 2041 2055: gap of unknown length
* 2056 10960: contig of 8905 bp in length
* 10961 10975: gap of unknown length
```

```
* 10976 21821: contig of 10846 bp in length
* 21822 21836: gap of unknown length
* 21837 30764: contig of 8928 bp in length
* 30765 30779: gap of unknown length
* 30780 37500: contig of 6721 bp in length
* 37501 37515: gap of unknown length
* 37516 47190: contig of 9675 bp in length
* 47191 57255: contig of 10050 bp in length
* 57256 57270: gap of unknown length
* 57271 68424: contig of 11154 bp in length
* 68425 68439: gap of unknown length
* 68440 79779: contig of 11339 bp in length
* 79780 93036: contig of 13243 bp in length
* 93037 93051: gap of unknown length
* 93052 96468: contig of 3417 bp in length.
FEATURES
Location/Qualifiers
1..96468
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/clone="Y74A11Y"
BASE COUNT 30222 a 17588 c 18100 g 30408 t 150 others
ORIGIN
Alignment Scores:
Pred. No.: 97.7 Length: 96468
Score: 116.00 Matches: 49
Percent Similarity: 39.09% Conservative: 28
Best Local Similarity: 24.87% Mismatches: 66
Query Match: 14.93% Indels: 54
DB: 2 Gaps: 9
US-09-847-539A-6 (1-159) x AC006902 (1-96468)
Qy 9 ArgIleIleProAsnGlyGlyThrLeuThrAsnLeuLeuGlyAsnAla----- 24
||||| ||||| |||||
Db 94762 CGCATCGCTCCACATCTCGCCACATTTGGTTGAAGCTGTCAACGATGTTCCAGCCAGTGTT 94703
Qy 25 ---ProGluLysLeuAlaLeuArg-
|||
:::|||||
Db 94702 GAGCCTTCAGCCGTCGCTTTTCGCGTGCCTGATGCTCCGCTTAATTCGTTCTGATCTTGAAGAAG 94643
Qy 32 -----AsnGluGluArgAlaIleAspGluLeuLysLysGlnAla--- 44
:::|||||
Db 94642 AACATCAAAAGACTGTTGACGATGAAAGCGCCGATGAACCTGAAGACGATGTCGGA 94583
Qy 45 -----IleGluAsp-----LysGluAlaThrAlaIle 54
:::|||||
Db 94582 AATCGGTCAAGAACGTCGGAAGATGTTGTTTCGAAATACCAAAATCAACCTCAACCACTC 94523
Qy 55 GluAlaAlaSerSerAspAla-----LeuGluAlaLeu 65
::: ||| |||||
Db 94522 GAGCTGCTAAGGACGACGCTTAACAACTCAAGCTACTGTAGACCACTTACCAAGCTG 94463
Qy 66 AlaAspGlnThrAspAlaLeuGlnSerGluGluAlaValValLysAlaAspAla 85
|||::: |||||
Db 94462 GCCGAGTCTACGACAAAGATCGATCCCAAGTCGCAAGGACATCAAGGACTCAAGACT 94403
Qy 86 AlaSer-----AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGln 101
::: ||| |||||
Db 94402 AAGCTTAAGGAATTCCTTCAGGCTCTTGAANAAGCCATTCACCAAGAGAGCGCTATTTCGT 94343
Qy 102 SerGluGluAlaGluValValGlnSerAspAlaAlaSerAspAlaTrpGluLysAla 121
|||::: |||||
Db 94342 CGTGAACAGCGGAGATT-----AATGATCGTCTCAACAACTCTCGAAGGAG 94295
Qy 122 AlaThrProIleAlaLeuAspValLysThrLysAspThrLysProValValLys 141
|||::: ||| |||
Db 94294 CTTACCAAGGTG-----GATGAGTTCAAACGAGAGATGCACCTCCCAATGTT---GAC 94244
Qy 142 GluGluArgGlnAsnValAsnThrLeuProThrThrGlyGluGluSerAsn 158
::: ||| |||||
```

SOURCE		Caenorhabditis elegans.	
ORGANISM	Caenorhabditis elegans		
REFERENCE	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.		
AUTHORS	Waterston,R.H.		
TITLE	The sequence of Caenorhabditis elegans clone Unpublished		
JOURNAL	2 (bases 1 to 44090)		
REFERENCE	Waterston,R.H.		
AUTHORS	Direct Submission		
TITLE	Submitted (23-FEB-1999) Genome Sequencing Center, Washington		
JOURNAL	University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA		
COMMENT	* NOTE: This is a 'working draft' sequence. It currently consists of 16 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.		
	1	2626:	contig of 2626 bp in length
	*	2627	2635: gap of unknown length
	*	2636	5780: contig of 3145 bp in length
	*	5781	5789: gap of unknown length
	*	5790	8469: contig of 2680 bp in length
	*	8470	8478: gap of unknown length
	*	8479	10474: contig of 1996 bp in length
	*	10475	10483: gap of unknown length
	*	10484	14663: contig of 4180 bp in length
	*	14664	14672: gap of unknown length
	*	14673	17678: contig of 3006 bp in length
	*	17679	17686: gap of unknown length
	*	17687	19794: contig of 2108 bp in length
	*	19795	19802: gap of unknown length
	*	19803	23618: contig of 3816 bp in length
	*	23619	23626: gap of unknown length
	*	23627	26716: contig of 3090 bp in length
	*	26717	26724: gap of unknown length
	*	26725	29127: contig of 2403 bp in length
	*	29128	29135: gap of unknown length
	*	29136	31536: contig of 2401 bp in length
	*	31537	31544: gap of unknown length
	*	31545	34131: contig of 2587 bp in length
	*	34132	34139: gap of unknown length
	*	34140	36477: contig of 2338 bp in length
	*	36478	36485: gap of unknown length
	*	36486	38730: contig of 2245 bp in length
	*	38731	38738: gap of unknown length
	*	38739	41003: contig of 2265 bp in length
	*	41004	41011: gap of unknown length
	*	41012	44090: contig of 3079 bp in length.
FEATURES	Location/Qualifiers		
source	1..44090		
	/organism="Caenorhabditis elegans"		
	/db_xref="taxon:6239"		
	/clone="Y48F5"		
BASE COUNT	13925 a	8467 c	8024 g 13485 t 189 others
ORIGIN			
Alignment Scores:			
Pred. No.:	43.7	Length:	44090
Score:	116.00	Matches:	49
Percent Similarity:	39.09%	Conservative:	28
Best Local Similarity:	24.87%	Mismatches:	66
Query Match:	14.93%	Indels:	54
DB:	2	Gaps:	9
US-09-847-539A-6 (1-159) x AC006784 (1-44090)			
QY	9	ArgilleProAsnGlyThrLeuThrAsnLeuLeuGlyAsnAla	----- 24



ANKLKATVQLTKLAESSDKIDPQVAKDIDKSKTKAKELLOALEKAIPQEDAIRREQA  
EINDNNLEKLETVDFEPEDALPIVDOLAANTNTLATDSDNNKAVPQESLISH  
DDLVGLPQFQLOHAIIDDKQALNKAAVNEIAPKLOLVSOQLSVQVPEVPSLISH  
OKQLLEDVENQKHNLLENLANIPENDPTADELROKSDWLSRLKDLKOLGSAVGDKL  
AALAFNAARKNAEDALLDITREDGGDNKSPDELIDDRGSTGVSNGVKLAAAFN  
AARKNAEDALLDITREDGGDNKSPDELIDDRGSTGVSNGVKLAAAFN  
AARKNAEDALLDITREDGGDNKSPDELIDDRGSTGVSNGVKLAAAFN  
RAEYNDLLARLATAADVLNKKRAELQAVAKADEKSLHSDVDRIVSRVPLVRESDE  
LRHNAEAVPTQAPAKAEELKEVEAAKAVIANAPSSDAHQVQOOLEQAVATAEFLIPLE  
DNKIANLQRISELLPLESAYADVRFFVDVDAEQTRHOYDNDVNLVNAELEDTELKQS  
ASQVANEIDDISKMDIDSTPERSILDTIAKSDIPALKAQINRIKIVNADSRKHVT  
TDPKTAEDLNKLAPEELDLAIKTSDEHREQLTSLKLNISQEPQLPDLQSDSD  
LKTAEKITSNKPAAEPPLAKIQELREAKRQVDEARSAHDIQVIALEKEDPQVAK  
ESAKKIDKKSPOEMIDELSAKVVAKALPKIPIEAAKNNELPAKDPKPAEFVSN  
LEAFVKDVTQSKQDELDKLNNAIDAIRKLGDAALDAEKTVPSSVPALSEFKDRI  
APHLATLVAENDVPASPEVALDRRAAKFVSQLEKNIQTKGDEKRADELKNDVG  
AVKVEDVYSKYQNPQPLDVAKODANKLKVATVOLTTLAESSDKIDPOVAKDKDS  
KTKAKELLOALEKAIPQEDAIRREQAETLIPDLERARLNWFLAARNDDIALTEQLQ  
AKAVIANAPSSDAHQVQOAVATAEFLIPDLERARLNWFLAARNDDIALTEQLQ  
PLDVAIQPKRGAEEAAQVENDLRNNSQSLDNKIANLQRISELLPLESAYADVR  
FDVDAEQTRHOYDNDVNAELEDTELKQASOVANEIDDISKMDIDSTPERSIL  
DITAKSDIPALKAQINRIKIVNADSRKHVTOPKIAEDLNKLAPEELDLAIK  
TSEHDKQELIISLNISQEPQLPDLQSDDLKTAETITNSUKPAEAPPLAKIQ  
ELREKRVGEARSAAHQDQVIALEKEDVAKESAKKKKKKKSPQEMIDELSAK  
VSAKALPIEAAKNNELPAKDPKPAEFVSNLEAFVKDVTQSKQDELDKLNNA  
NDAIKRIGDALDAEKTVPSSVPALSEFKDRIAPHLATLVAENDVPASPEVAL  
RDRAKFVSDLEKNTQKTGDEKRADELKNDVGNAKVNVEDVYSKYQNPQPLDVAK  
DANKLATVQELTKLAESSDKIDPQVAKDIDKSKTKAKELLOALEKAIPQEDAIRREQA  
AEINRLNLEKLETVDFEPEDALPIVDOLAANTNTLATDSDNNKAVPQESLISH  
HDDLVLGPLEKVFQLOHAIIDDKQALNKAAVNEIAPKLOLVSOQLSVQVPEVPSLISH  
EQQLLEDVENQKHNLLENLANIPENDPTADELROKSDWLSRLKDLKOLGSAVGDKL  
LAALAFNAARKNAEDALLDITREDGGDNKSPDELIDDRGSTGVSNGVKLAAAFN  
PDELDDKRAEYNDLLARLATAADVLNKKRAELQAVAKADEKSLHSDVDRIVSRVPLVRESDE  
PLVRESDELHNAEAVPTQAPAKAEELKEVEAAKAVIANAPSSDAHQVQOOLEQAVATAEFLIPLE  
FLYIPDLERARLNWFLAARNDDIALTEQLQPLDVAIQPKRGAEEAAQVENDLRNNSQSLDNKIANLQRISELLPLESAYADVR  
DETLLKQASOVANEIDDISKMDIDSTPERSILDTIAKSDIPALKAQINRIKIVNADSRKHVT

## Alignment Scores:

Pred. No.: 45, 2 Length: 54126  
Score: 117.00 Matches: 49  
Percent Similarity: 39.09% Conservative: 28  
Best Local Similarity: 24.87% Mismatches: 66  
Query Match: 15.06% Indels: 54  
DB: 3 Gaps: 9

US-09-847-539A-6 (1-159) x AC006834 (1-54126)

Qy 9 ArgileleProAsnGlyThrLeuThrAsnLeuGlyAsnAla----- 24  
Db 22833 CGCATCTCCACATCTCGGCATTTGGTTGAAGCTGTCACGATGTTCCAGCGAGTGT 22892  
Qy 25 ---ProGluLysLeuAlaLeuArg----- 31  
Db 22893 GAGCCTTCAGCGCTGTTGGTTCGTCGCGCTAAATTCGTTCTGATCTTGAAG 22952  
Qy 32 -----AsnGluGluAlaGluAlaIleAspGluLeuLysLysGlnAla--- 44  
Db 22953 AACATCCAAAGACTGCTGACGATGAAGACCGCGCGCATGAAAGAACGATGTCGGA 23012  
Qy 45 -----LysGluAsp-----LysGluAlaThrAlaIle 54  
Db 23013 AATCGCGTCAAGACGTCGGAAGATGTTGTTTCGAATATACCAAAATCAACCTCAACCACTC 23072  
Qy 55 GluAlaAlaSerSerAspAla-----LeuGluAlaLeu 65  
Db 23073 GACGTGCGTAAAGACGACGCTAAACAACTCAAGGCTACTGTAGAGCAACTTACCAAGCTG 23132

Qy 66 AlaAspGlnThrAspAlaLeuGlnSerGluGluAlaValValLysAlaAspAsnAla 85  
Db 23133 GCCGAGTCTAGCGACAAAGATCGATCCACAAAGTCCCAAGGACATCAAGACT 23192  
Qy 86 AlaSer-----AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGln 101  
Db 23193 AAGGCTAAGGAATGCTTCAGGCTCTTGAAGAAAGCCATTCACAAAGGAGCGCTATTCTG 23252  
Qy 102 SerGluGluAlaGluValValGlnSerAspAlaAlaSerAspAlaTrpGluLysAla 121  
Db 23253 CGTGAACAAAGCCGAGATT-----AATGATCGTCTCAACAAAGCTCGCAAGAGAG 23300  
Qy 122 AlaThrProIleAlaLeuAspValLysThrLysAspThrLysProValValLysLys 141  
Db 23301 CTTACCAAGGTG-----GATGACTTCAACACAGGAGTGCATCTTCAATGTT--GAC 23351  
Qy 142 GluGluArgGlnAsnValAsnThrLeuProThrThrGlyGluGluSerAsn 158  
Db 23352 CAACTCGCGCAACTACTAACAACCTTGAAGACTGCCACTGATTCGAACAAC 23402  
RESULT 27  
AL604063/c 146759 bp DNA linear ROD 29-NOV-2001  
LOCUS Mouse DNA sequence from clone RP23-467J12 on chromosome 11,  
DEFINITION complete sequence.  
ACCESSION AL604063 GI:17721230  
VERSION AL604063.4  
KEYWORDS HTG.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 146759)  
AUTHORS Lovell, J.  
TITLE Direct Submission  
JOURNAL Submitted (29-NOV-2001) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
humquery@sanger.ac.uk  
COMMENT On Nov 30, 2001 this sequence version replaced g1:1644787.  
During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.  
This sequence was finished as follows unless otherwise noted: all  
regions were either double-stranded or sequenced with an alternate  
chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by at least  
one plasmid subclone or more than one M13 subclone; and the  
assembly was confirmed by restriction digest. The following  
abbreviations are used to associate primary accession numbers given  
in the feature table with their source databases: Em: EMBL; Sw:  
SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information on the WORMPEP  
database can be found at  
http://www.sanger.ac.uk/projects/C\_elegans/wormpep RP23-467J12 is  
from the RP23-467J12 Mouse PAC Library  
constructed by the group of Pieter de Jong.  
For further details see http://www.chori.org/bacpac/home.htm  
VECTOR: pBACe3.6  
IMPORTANT: This sequence is not the entire insert of clone  
RP23-467J12. It may be shorter because we sequence overlapping  
sections only once, except for a short overlap.  
The true left end of clone RP23-366M19 is at 144760 in this  
sequence. The true right end of clone RP23-259D24 is at 2000 in  
this sequence.  
FEATURES  
source Location/Qualifiers  
1..146759  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/chromosome="11"  
/clone="RP23-467J12"



```
investigating biology. The C. elegans Sequencing Consortium
Science 282 (5396), 2012-2018 (1998)
99069613
2 (bases 1 to 54126)
Gattung,S., Goeta,D. and Broym,M.
The sequence of C. elegans cosmid zk973
Unpublished
3 (bases 1 to 54126)
Waterston,R.
Direct Submission
Unpublished
4 (bases 1 to 54126)
Waterston,R.H.
Direct Submission
Submitted (23-FEB-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
5 (bases 1 to 54126)
Waterston,R.H.
Direct Submission
Submitted (01-MAR-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
6 (bases 1 to 54126)
Waterston,R.
Direct Submission
Submitted (08-MAR-2000) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
7 (bases 1 to 54126)
Waterston,R.
Direct Submission
Submitted (28-MAR-2000) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
8 (bases 1 to 54126)
Waterston,R.
Direct Submission
Submitted (03-OCT-2001) Department of Genetics, Washington
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
Louis, MO 63110, USA
On Oct 5, 2001 this sequence version replaced gi:4263104.
Submitted by:
Genome Sequencing Center
Department of Genetics, Washington University
St. Louis , MO 63110, USA, and
Sanger Centre, Hinxton Hall
Cambridge CB10 1RQ, England
email: rwenematode.wustl.edu and jes@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone.
It may be shorter because we only sequence overlapping sections
once, or longer because we provide a small overlap between
neighboring submissions.

This sequence was finished as follows unless otherwise noted: all
regions were double stranded, sequenced with an alternate chemistry
or covered by high quality data (i.e., phred quality >= 30); an
attempt was made to resolve all sequencing problems, such as
compressions and repeats; all regions were covered by sequence from
more than one ml3 subclone.

NOTES:
Coding sequences below are predicted from computer analysis, using
the program GeneFinder(P. Green and L. Hillier, ms in preparation).
FEATURES             source
1..54126
/organism="Caenorhabditis elegans"
/strain="Bristol N2"
/db_xref="taxon:6239"
/chromosome="1"
/clone="zk973"
11..25772
/gene="zk973.6"

source
gene

join(11..103,170..209,1232..1563,1609..1736,1784..1940,
1993..2127,2196..3029,3082..3300,3352..4719,4767..5015,
5065..9450,9629..12490,12669..15377,15556..18264,
18443..21151,21330..24038,24217..24654,24700..24843,
24890..25021,25098..25586,25635..25772)
/gene="zk973.6"
/note="contains similarity to myosin-like proteins; coded
for by the following C. elegans cDNAs: cm5b12, cm10f6,
cm15g2, yk2f4.3, yk2f4.5, yk7e10.3, yk7e10.5, yk19b4.3,
yk19b4.5, yk32c4.3, yk32c4.5, yk36b3.3, yk36b3.5,
CEES066F, CEMSA37F, CEMSH10F, yk25a11.5, yk8f4.5,
yk752h02.5, yk871h03.5, yk54d5.5, yk855c06.5, yk44g3.5,
yk539e4.5, yk542g12.5, yk369e5.5, yk881h02.3, yk539e4.3"
/codon_start=1
/product="Hypothetical protein zk973.6"
/protein_id="AAF40010.2"
/db_xref="gi:15967153"
/translation="MELFRLRQFCDAVKILRSQNTKWNKIKISQCCRRGGGEYIIGS
ECGFAEAFSGKFNLSLVSQCFQIYLCFLSQQLKLVISRVLFSLKLTTCGCPLLF
SSTCRSPSLTLMSLFORSKSAIYHTRFLFSORRGLODMANRTRTHLKHESKSIDT
LIVRLDEMTODANOLRAELASIHGSPFQLEFDRLEKLNMLIHDKDKRLMLVHRRH
YLTAANLITDSKTDLAQRATSSSDHPDILRATDEVTAKLDKEELRRLGAEAMNI
TDLVVVAILFSRRRQSGEPEVEELRALRELIARTEPRDPSDAIVADILRM
KDEKROKTTIDEQTATLTDEQRASFAPLIEDYRRADRHRIYFEHLVMTLSLAE
QFLEDEIGMTIQTSRADLLRNMSTENKWKTDLANIERQVQDPEPERSYRFGSL
HRKDSMEARINKYLTSAKINAKLTQPEKWLNAIEDEIQETQDPEPERSYRFGSL
HEVALAKORLVAKLERLVANKEEVLCERYHTIMKLTPTQTAGVPLHVSYNLDR
NGPQSQISVSSIASSELERPESVMSLTSSIGVTPADVAELSPPEAKINKLLOKHII
EDSLTKGPKPIDTVREDDVKLEKTRNRAEILQQLSTNSIEDAEKGGUKHFLMLNN
YDDLSTENIEDRDNELTAKNQEILAELSNAEQTLQNSPLEDLDIAEFLRLQMLND
LVKVMCNPKRYVECELIDSSREGSPQRRRRKKVMVMSNTITTIHVVEERLEAS
DISNPQVQKLVAVKENLRDLTITVTPHPPSIMPGLCTNRNDEEVKLAELADR
AIDTASSMVEDAPTEDEAKSAIHLDDQKVTLNHLHVLEGIEPKNQDQRTDADIA
SVSGEKLNLKSAVEEVYEVYLASTNPVKEDQLOHQIEQVOTPAEQSNMDDPSRQ
PPVLSDERIELKTDPSAIDKEVRSDESDPVKIIDLFGQLQTAVNEASPLACETDDV
DALQVADSKLTQDRTIRKIHAILDITDDPVQKPAIIESDKIKDINNARDNNRQI
DNLYNQTVPVVPKSTPLNEIDAVRQASTVYSDBLCTNCKILLSARQTLANRQKQ
VDSVEAVVNCNPTETVPTILEEYTLVDNIERKLANEIEVPTNDRQDLVQLOQLD
VILEEVEVNVNCKTEKESKLEKPLDLOIGDNVKEKLSRQSDPSDAIDALS
NVHQYNATINSIDDKIDELKNPDEDTSAADQLISELHVISEMPAVTIDLSMLNAIEE
GLSTPLPQHAENVQAKIDELQKKEVADQTEQILSDNLAFGDMFAITLIDLKLSVED
GIATLVPEDSERIKAKIVDLRRKKEDQQAELQLOELSVISDMPLSTLDLNNLQIED
NLNSLPAAEDAKIREKINLRRRKOESDQAEALQLOELSVISDMPLSTLDLNNLQIED
NLNSLPAAEDAKIREKINLRRRKOESDQAEALQLOELSVISDMPLSTLDLNNLQIED
NLNSLPAAEDSKIREKINLRRRKOESDQAEALQLOELSVISDMPLSTLDLNNLQIED
NLNSLPAAEDSKIREKINLRRRKOESDQAEALQLOELSVISDMPLSTLDLNNLQIED
GLSTLLSEDSKIQQAIDSRLKKSDSLAQHALEALSQSKLPSVSNLBELKLEE
TLSTVPEDSKVIKDKTAEALAKADHAENIYVELKKIEMDPIASGVSDVLATIED
QILQMPVOYQPSVKETLDKLOAKEEDDKLAGVYDELEKIAKLPAQDYNKLLAKIDE
KLNSLPKQDIAETHRKVEDIKVTRADIVQADVLVDKLPKADIDEHLNLSIEKPTIP
SDSSDQIAIGLKLRRKOANIDEGKTLNLEAEIOKMPADSLNEHALNLATSDKIF
GSEISDKTMOEIDVLIRKQNNHEVARLNAESVLOQLDKISEEPHLSLTERLAPFLQ
IDTVPACFVDKIRNKINEVKLHDEAVODEKELKELVAKVQNIKGTSIDVAVSDF
BEIEREINGSLEAEPEPLAKIOELREAKRVGDEARSAHQIIVALEKEAEDVTAKE
SAKKRDKKKSPQEMIDELSAVNAEKALPIKEEAANKENLPADDPKPAEQVSNL
EAFVKVDVQVSEKODELKNNAEKIRGLGDALDAAETKVVSPSPALSEFKDRIA
PHLATLVAVNDVPASVPSVALDRAAKFVSDLEKNIQKTDEKRADELKNDVNG
AVKVVEDVSVYQVQPOPLDVAKDDANKATVEOLTKLAESSDKIDQVAKDKDVK
TKAKELQALEKAIPOEDAIRREQAEINRLNNLEKELTVDEKPEKDALPVDQALAA
NTNLTKATDSNNNEKAVAPSSLSIHDDLVLVGLPEKVFQLOLQHAIDDKKQALNKAAYNE
LAPKQLVSVQQLQSVQPEVPAISLDEQKOLLEDVENKHNLENLANLPENDPTADELR
OKSOWDLSRLKDLKLGSAVGDKLAALAAFAFNAARKNAEDALLDITREDGDNNKSPD
ELIDDLAKKEETVAKLIDTVSVKVPDELLDKERAEYNLLARLATAAADLVAKNRAELE
QAVKAKADEKSLHDSVDRIVSVRLVPRESDELRNAEAVPTQVAPKAEELKVEEA
KAVTANAPRSSAAHVQQLQEQAVATAETLIPDLERARNLNEFLARNDELAIDLEQLOQP
LDVLAQPKRSAAEQVLENLRNNSQOLSOLDNRIANLRISELLDPLESAIDVRF
FDVDAEQTRHQYDDVLDNDVAAELEDETLKQSAQVANEIDDIKMTDPTDPERSIDL
TIAKSDIPALKAQINIRIDRVNADASRKHVTTDPKTAEDLDNKLAKIQTDLDDAIIKT
SDEHDKQOLILSLKNIISOFEIPLDQKSDOLKTAETKINSLKPEAEELKIAKIOE
LREAKRGDEARSAAHQIIVALEKEAEDVTAKESAKKKKDKKKKSPQEMIDELSAKV
EAKALIPKIEEAANKENLPADDPKPAEQVLSNLEAFVKDVTQVSEKODELKNNAAN
DAIKRLGDALDAAETKVVSPVPALEFKDRIAPHLATLVAVNDVPASVPSVALR
DRAAKFVSDLEKNIQKTGDDDEKRADELKNDGNAVKNVYEDVSVYQVQPOPLDVAKDD
```

Yates,J.L., Camiolo,S.M., Ali,S. and Ying,A.  
Comparison of the EBNA1 proteins of Epstein-Barr virus and  
herpesvirus papio in sequence and function  
Virology 222 (1), 1-13 (1996)  
96400106  
806482  
2 (bases 1 to 3222)  
Direct Submission  
Submitted (31-MAR-1995) John L. Yates, Department of Human  
Genetics, Roswell Park Cancer Institute, Elm and Carlton Streets,  
Buffalo, NY 14263, USA  
Location/Qualifiers  
1. .3222  
/organism="Herpesvirus papio"  
/strain="594-S"  
/specific\_host="Papio sp."  
/db\_xref="taxon:10394"  
/clone="p701 (5.4 kbp SalI to EcoRI HVP DNA insert)"  
/lab\_host="Papio sp."  
<1. .294  
/note="similar to BRRF2 of Epstein-Barr virus, Swiss-Prot  
Accession Number P03210"  
/codon\_start=1  
/protein\_id="AAA66372.1"  
/db\_xref="GI:775214"  
/translation="GDIEDVSEARASTPLPPGDEIEFEAWLESONPSLEDVQOEF  
AGMRTVEAEEDNSDDCEFSDDWASDSDHEDGEGTSGGSRSLHSLYSV"  
330. 1760  
/function="binds to oriP to permit replication of the  
viral episome"  
/note="similar to EBNA1 of Epstein-Barr virus, Swiss-Prot  
Accession Number P03211"  
/citation=[1]  
/codon\_start=1  
/evidence="experimental"  
/product="EBNA1"  
/protein\_id="AAA66373.1"  
/db\_xref="GI:775215"  
/translation="MSDEGPGNNGLGEKGTGGGTGRGGHGRGRGRGRGRG  
GSRGLAGTGGSGGTGLDGLGPGPPNKKRSCVCGKGGSGARGTGGSGAGAG  
GSGAGAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG  
RGRGRGGSGSGR  
TRASGSSGSSGSPVFPCHNSAPLTPATPLGGDRGTPDPGDEPPGAMGQPPDP  
GEGPSHRPPGPGPGKGMFGVRRGGYGSKEKMAQSLRLLSRVCOVPTNPE  
GDWPAVMVYGPKNCSYLRARLCGCCVPMCRPLPSRLPYGHSWGTGPEPTPLMESCV  
SYFLVFLPTQGSARECYDALYDISTRPQTSSVKVTFCTDPPVPLPIFYPPPEPT  
GSGAGGEGAGGDDGNGSDGEGEGE"  
612. .770  
repeat\_region  
/note="The 21-bp repeat unit encodes the heptapeptide  
Gly-Gly-Ser-Gly-Ala-Ala; approximately seven and  
one-half copies of this repeat in strain 594-S encode a 53  
amino acid repetitive domain."  
/citation=[1]  
/function="encodes heptamer repeat of Gly, Ser and Ala"  
/rpt\_family="analogous to IR3 of EBV"  
/rpt\_type=tandem  
/rpt\_unit=612. .622  
612. .770  
/gene="EBNA1"  
1843. .2259  
CDS  
/note="potential 15kDa protein; similar to BRRF2 of EBV,  
Swiss-Prot Accession Number P03212"  
/codon\_start=1  
/protein\_id="AAA66374.1"  
/db\_xref="GI:775216"  
/translation="MHAARIVFVSLVALFLVLTGWNWAPCPCHVTLSQQHHLALE  
NISDIYVSKNTCDGFSLAGNSKNSGNSOLVSRCANGLNVAIFYIIVLRSSALTS  
SOLLELNTLESILYASVSDVDFCANLRYAWHHAG"  
2241. 3008  
/note="similar to BRRF3 of EBV encoded by GenBank  
Accession Number V01555; similar to EBV ucrail DNA  
glycosylase, Swiss-Prot Accession Number P12888"  
CDS

Yates,J.L., Camiolo,S.M., Ali,S. and Ying,A.  
Comparison of the EBNA1 proteins of Epstein-Barr virus and  
herpesvirus papio in sequence and function  
Virology 222 (1), 1-13 (1996)  
96400106  
806482  
2 (bases 1 to 3222)  
Direct Submission  
Submitted (31-MAR-1995) John L. Yates, Department of Human  
Genetics, Roswell Park Cancer Institute, Elm and Carlton Streets,  
Buffalo, NY 14263, USA  
Location/Qualifiers  
1. .3222  
/organism="Herpesvirus papio"  
/strain="594-S"  
/specific\_host="Papio sp."  
/db\_xref="taxon:10394"  
/clone="p701 (5.4 kbp SalI to EcoRI HVP DNA insert)"  
/lab\_host="Papio sp."  
<1. .294  
/note="similar to BRRF2 of Epstein-Barr virus, Swiss-Prot  
Accession Number P03210"  
/codon\_start=1  
/protein\_id="AAA66372.1"  
/db\_xref="GI:775214"  
/translation="GDIEDVSEARASTPLPPGDEIEFEAWLESONPSLEDVQOEF  
AGMRTVEAEEDNSDDCEFSDDWASDSDHEDGEGTSGGSRSLHSLYSV"  
330. 1760  
/function="binds to oriP to permit replication of the  
viral episome"  
/note="similar to EBNA1 of Epstein-Barr virus, Swiss-Prot  
Accession Number P03211"  
/citation=[1]  
/codon\_start=1  
/evidence="experimental"  
/product="EBNA1"  
/protein\_id="AAA66373.1"  
/db\_xref="GI:775215"  
/translation="MSDEGPGNNGLGEKGTGGGTGRGGHGRGRGRGRGRG  
GSRGLAGTGGSGGTGLDGLGPGPPNKKRSCVCGKGGSGARGTGGSGAGAG  
GSGAGAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG  
RGRGRGGSGSGR  
TRASGSSGSSGSPVFPCHNSAPLTPATPLGGDRGTPDPGDEPPGAMGQPPDP  
GEGPSHRPPGPGPGKGMFGVRRGGYGSKEKMAQSLRLLSRVCOVPTNPE  
GDWPAVMVYGPKNCSYLRARLCGCCVPMCRPLPSRLPYGHSWGTGPEPTPLMESCV  
SYFLVFLPTQGSARECYDALYDISTRPQTSSVKVTFCTDPPVPLPIFYPPPEPT  
GSGAGGEGAGGDDGNGSDGEGEGE"  
612. .770  
repeat\_region  
/note="The 21-bp repeat unit encodes the heptapeptide  
Gly-Gly-Ser-Gly-Ala-Ala; approximately seven and  
one-half copies of this repeat in strain 594-S encode a 53  
amino acid repetitive domain."  
/citation=[1]  
/function="encodes heptamer repeat of Gly, Ser and Ala"  
/rpt\_family="analogous to IR3 of EBV"  
/rpt\_type=tandem  
/rpt\_unit=612. .622  
612. .770  
/gene="EBNA1"  
1843. .2259  
CDS  
/note="potential 15kDa protein; similar to BRRF2 of EBV,  
Swiss-Prot Accession Number P03212"  
/codon\_start=1  
/protein\_id="AAA66374.1"  
/db\_xref="GI:775216"  
/translation="MHAARIVFVSLVALFLVLTGWNWAPCPCHVTLSQQHHLALE  
NISDIYVSKNTCDGFSLAGNSKNSGNSOLVSRCANGLNVAIFYIIVLRSSALTS  
SOLLELNTLESILYASVSDVDFCANLRYAWHHAG"  
2241. 3008  
/note="similar to BRRF3 of EBV encoded by GenBank  
Accession Number V01555; similar to EBV ucrail DNA  
glycosylase, Swiss-Prot Accession Number P12888"  
CDS

```
CVSGSVACKMLTTLGLTIDDVSDGSOANAGQRLIRAQGNAAALHQGWKAIONVLEW
SCFNSDYLDQMLALLRYPDGNCTACTRKAIPGTRLLASQAILAAGGIHRRAGCAETI
LLFNHRHFLHNDNDWGAIINQLRHAPEKKIVVEADTPKEAIAALRAQDPVQLQDKF
SPOOATEIAQIAPSLAPHCTLALTGGINLTTLKNVLDGIRLFITSAPYYAAPADIKV
SLOPAAS1"
complement(3219. .3500)
/gene="Ecs1695"
complement(3219. .3500)
/gene="Ecs1695"
/note="unknown"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB35118.1"
/db_xref="GI:13361160"
/transl_table="MPTLILSVDKIANRITAPRNVLSRSTSGVLARLTMSVSGYIAG
INRMKLVSPPLPAATGRSSGGIAYRRHCDDEPFGNRCVTSGRYGR"
complement(3445. .4089)
/gene="Ecs1696"
complement(3445. .4089)
/gene="Ecs1696"
/note="unknown, similar to hypothetical proteins e.g. Orf2
[Escherichia coli CFT073] gi|3661479|gb|AAC61712.1|
percent identity 98 in 214 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB35119.1"
/db_xref="GI:13361161"
/transl_table="MLDIDDIFADLYLQOLKLHRTKTPDHWQRAEKMAENCASPT
DSYQLQILAKIDQAGTDFDMGCGPGTVSLALADKLTIIYGVDSOGLNVAARRAA
ALKANDVHILORAEEDSDLPDGDIAVSRSTLVADMQRQMSKLNNQARLVRTTHL
VSTSPVPAIORAGREVIEPNVIFALNVLYQMGYAHVDFIRQCNCQDQNSI"
complement(4099. .4857)
/gene="Ecs1697"
complement(4099. .4857)
/gene="Ecs1697"
/note="probable ferric enterobactin transport, similar to
ferric enterobactin transport ATP-binding protein
[Escherichia coli CFT073] gi|3661480|gb|AAC61712.1|
percent identity 79 in 148 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative ferric enterobactin transport
ATP-binding protein"
/protein_id="BAB35120.1"
/db_xref="GI:13361162"
/transl_table="MTLIVVRHSLGYSRHPVLRDVSFTLSQGTICLLGANGCGKTT
LMRSILGVIPLLLKGEVLDSLVPQVTLSHRQRAQAIWVPAQHDGFAPSVLDMVLMGL
APTIGAFVPGKQERLKAETEOLKILHLAARWNTLSGGEROLALATARALVOOPRL
LLIDEPASSLDGFIQIOLDDTLAOLKNGMTMLSTHTPLHANAIDSIIQVDPGRV
TQGLPTEQLTNKLAALRVSAADQIHHLHLSAISH"
complement(4854. .5834)
/gene="Ecs1698"
complement(4854. .5834)
/gene="Ecs1698"
/note="probable ABC transporter protein (permease), similar
to ABC transporter permeases e.g. [Haemophilus influenzae]
gi|2501391|sp|Q57130|YE71_HAEIN percent identity 40 in 323
aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative ABC transporter permease protein"
/protein_id="BAB35121.1"
/db_xref="GI:13361163"
/transl_table="MNVIGCTILLAAISITFAAVSGAYHLDQMOQLALILROENVPV
QEOIVFMDIVPRILAAALLGAAGACTTYQGLRNPLVSPDILGVSGAGLIGACTA
ILMGLSIYVLIQYAFCGGLMVVAGVWLITRVRTRHPDILTILVLGVLAGTLCGAGISL
IKTLADPTIQTUPSITFWLLGGSLSTVTLRDICLYAAPILIGLSLPFFFLWRMNLTLSD
```

```
DEARSLGNVTRLRFLVILVCATLITASTVAITAGIIGWGLVPHPIARLLTGHNHQQLL
PMACHTGAILLLLTDLTARSIGTTEIPILGITAFVGAPFFLLLLLRGGR"
complement(5834. .6856)
/gene="Ecs1699"
complement(5834. .6856)
/gene="Ecs1699"
/note="probable ABC transporter, weakly similar to iron
(iii) ABC transporter, ATP-binding protein [Pyrococcus]

Alignment Scores:
Pred. No.: 164 Length: 291136
Score: 119.50 Matches: 50
Percent Similarity: 38.59% Conservative: 21
Best Local Similarity: 27.17% Mismatches: 66
Query Match: 15.38% Indels: 47
DB: 1 Gaps: 5

US-09-847-539a-6 (1-159) x AP002556 (1-291136)
QY 15 GlyThrLeuThrAsnLeuLeuGlyAsnAlaProGlu-----LysLeuAlaLeu 30
||||| :|: ||||| |||
Db 111736 GGTACGCTGAATGATTTCTCGTGCCATGTCGGAGGATGACGTCCGGCCGGAGGCACTG 111795
QY 31 ArgAsnGluGluArgAlaIleAspGluLeuLysLysGlnAlaIleGluAspLysGlu--- 49
||| ||| :|: ||||| :|: ||| ||| |||:
Db 111796 CGTCGTTTGAACCTGATGTTGGAAGAAGCGCGCTCAGCTGAGGAGCGGAAGAAT 111855
QY 49 ----- 49
Db 111856 GCCGAGAGCGGAGACGTCCCGGAGGAATGCCGGCATATCAGCAGTCAGCGCAGAAGAG 111915
QY 50 -----AlaThrThrAlaIleGluAlaIleSerSerAspAlaLeuGluAlaLeu 65
||| |||: ||||| ||| :|: |||
Db 111916 AACGCTGCAATGCTGACACTTCAGCAGGGGATGCATCGAGTCAGCCGGCAGCGGCA 111975
QY 66 AlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLysAlaAspAsnAla 85
||| ||| ||||| |||||: ||| |||
Db 111976 GAAAGTGACGCGCTGCAAGCAGTCAGAGAGGGGTCTCTGCTCGGCTCTCGGCGC 112035
QY 86 AlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAla 105
||| ||| |||: ||| ||||| ||| :|: |||
Db 112036 GCTCAAAAGCCAGTGAGTCATTACAAGTGCACAGATGCTGAGTTGTCAAAAAGACG 112095
QY 106 GluValValGlnSerAspAsnAlaAlaSerAspAlaTrpGluLysAlaAlaThrProile 125
:|: ||||| ||||| |||
Db 112096 GCAGAAAGTGCAGCGCGTAAATGCAGCCAGGGATGCAACAGCCGCGACGA----- 112143
QY 126 AlaLeuAspValLysLysThrLysAspThrLysProValValLysLysGluGluArg--- 144
:|: ||| :|: |||: |||
Db 112144 -----GAAAAAGCCCGGAGTCAAGCAGAAAGCGCACAGTCAGCGCAACAAAGC 112191
QY 145 -----GlnAsnValAsnThrLeuProThrThr-----GlyGlu 155
:|: ||||| ||||| |||
Db 112192 AGCATAGCGCGGAGAGACCGCTAAACCGATCCCCACCGTGGTGGGCGCTCCCGGGCCA 112251
QY 156 GluSerAsnPro 159
:|: |||
Db 112252 AAGGGGGAACCG 112263

RESULT 25
HPU23857
LOCUS
DEFINITION
Herpesvirus papio BRRF2 homolog gene, partial cds, EBNAL1, BRRF2
homolog and BRRF3 homolog genes, complete cds, and BRRF4 homolog
gene, partial cds.
ACCESSION
U23857
VERSION
U23857.1
KEYWORDS
GI:775213
SOURCE
Herpesvirus papio.
ORGANISM
Herpesvirus papio
Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
Gammaherpesvirinae; Lymphocryptovirus.
REFERENCE
1 (sites)
```



```

/transl_table=11
/evidence-not_experimental
/product="putative polyprotein"
/protein_id="BAB34846.1"
/db_xref="GI:13360887"
/translacion="MKTRHHVQGDITKLVADVIVNAANSLMGGGGVDAIHRAGCP
ALLDACLVKQQQDCPTGHAVITLAGDLPAKAVVHTVGPWVRGGEQNEOLLQDAYL
NSLRVAANSYVAFPAISITGVYGPRAAAAEIAVTVSEFITHRALPEQVYFCVD
EENAHLYERILLTQQGDE"
549..2030
/gene="ECS1424"
549..2030
/gene="ECS1424"
/note="similar to YMDC_ECOLI gi|1787284 percent identity
99 in 493 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="putative synthase"
/protein_id="BAB34847.1"
/db_xref="GI:13360888"
/translacion="MMKKTPTSTKDSLPNKEMNDLPRLASAVPLCSQHQGCGLPFL
EKSIDAAARYLAEMSEHTLDVQYIYQDDMSGRLLFSALLAAAKRGVRRLLIDN
NTPGDLILLDDGRIEVRLEPNFSLRLLPUGYITDFSLRNRHNSKSFYVDGVV
TLVGRNIGDAIFGAGEEPLFSLDJYMAIGPVVEDVADDFAIRYWKSVSPLOQVLDV
PEGEMADRIELPASWHNDAMTHRYLRKMESPFIINHVDGTLPILMAKTRLLSDPAK
GEGKAKRHSILPQRLFDITMGSPSERIDIISSEYFVPTAGVAOLIRMVKRGKTAITLN
SLAANDVAVHAGVWRKLLRYGVVELYELKPTREOSSTLHDGRITGNSGASLHAKT
FSIDGKTVFISGFNFDPRSTLNTFMGFVIESETLAQLIDKRFIQSQYDAWQLRLDR
WGRINWDRHAKKEIVLKKEPATSFWRVVRVRLASILPVEWLL"
complement(2038..3195)
/gene="ECS1425"
complement(2038..3195)
/gene="ECS1425"
/note="similar to MDOC_ECOLI gi|3025226sp|P75920 percent
identity 99 in 385 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="glucans biosynthesis protein"
/protein_id="BAB34848.1"
/db_xref="GI:13360889"
/translacion="MNPVPAQREYFLDSIRAWMLLGLPFHISLIYSHTWHVNSAEP
SLWLTNDFTHSPRMQVFFVIGSYFMFLFRLPKKMMKVRVERVIGPMLTAIPLL
TLPOFILQYVKGAEKSPGLSLVDKYNLAWELISHLWLLVLVMTLCVFMFKRI
RNNLENDKTNKFSWKLSVIRICLIGIGVAVIRTFIVYPTLSNGTFNFIWQTL
FYLFFILGALAFIPHLKALFTTPSRGCTLAALAFVALLNORYSGDAWYETES
VITVGLWMVNVVFSFGHRLNFQASRVTFYNASLFYLVHHPHLFFGAYITPHI
TSNWLGLCGIFVVGITAILYEIHLRIPLLKFLFSKPVVYKRENDKAPAR"
3588..5123
/gene="ECS1426"
3588..5123
/gene="ECS1426"
/note="similar to MDQG_ECOLI gi|1787286 percent identity
99 in 511 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="periplasmic glucans biosynthesis protein"
/protein_id="BAB34849.1"
/db_xref="GI:13360890"
/translacion="MMKMRWLSAAVMLTLYTSSWAFSIDDVAQQAQSLAGKGYEAPK
SNLPSVFRDMKYAQQTQFNHDKYWNKLTPEKLEFYHOGMYFDTVPKINEYTATA
VKRIYSPDYFTFGVDHDKTDVLDLGFAGPKVLYPINSKDKNDIYVSLMGLASVRFVI
CAGQVYGLSAGLADLTALPSGEPEPEKPIERPKPTDKRLIYALLDSPRATGAY
KRYVMPGRDSTVVDQSKIIYLRDKVKGLKVAFLPSMFLFPNQPSPANNIRPELHDSNG
LSIHAGNGEWRSLNPNKPLHVASFSFSENPGGFLQGRDFFSFEEDLDDYDRSPS
ASWTPKGEWGSVELVEITNENDNINVAIYWTPDQLPEPKENKFKYITTFSRDED
KLHAPDNAWQVQTRSRSDGVKQSNLIRQPDGTIAFVVDFTGAEMKKLPEDPVPAOTS
IGDNGEIVESTVRYNPVTKGWRLVMRVKVKDKAKKTETEMRAALVNAQDTLSETWSYQLP
ANE"
5086..7659
/gene="ECS1427"

```

```

CDS
5086..7659
/gene="ECS1427"
/note="similar to membrane glycosyltransferase MDQH_ECOLI
gi|1787287 percent identity 100 in 847 aa but differs at
N-ter (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="membrane glycosyltransferase"
/protein_id="BAB34850.1"
/db_xref="GI:13360891"
/translacion="MRPGATSYLPMMKTTYIDAMPFAASEKALPKTDIRAVHQALD
AERTWAREDSDPGSVKARLEQAWPDSLADGQLIKDDGRDQLKAMPEAKRSMEFPD
PWTNPVGRFWRDLRDVTPRYLARLTKEEQSEQKRWTVGTRIRYILLILLAQTV
VATVYKRTILPYQWALINPDMVMGQDLVSMFMQLPYMLQGTGLILFAVLPQWVSAG
FWTALMGFLOLLIGROKYSISASTVGEPLNPEHRTALIMPICNEOVNRFVAGLRATW
ESVKATGNAKHEDVYITLSDSYNDPICVAEQKAMMELIAEVGGEQIYFRRRRRVRK
SGNIDPCRRWGSQSYTMVVDADSDVMTGQCLGLVKLMEANMAGLIQSSPKASMD
TLYARCOQFATRYVGLPTAGLHFWQLGESHYGHNAIIRVKPFIIEHCALAPLPGEGS
FAGSILSHDFEAAALRRAGWGYIAYDLPGSYEELPNLLDELKRDRRWCHGNLMNF
RLFLVGMHPVHRAVFLTGVMSYLSAPLWFMFLASTALQVHALTPQVYFLOPROLF
PVWPMPELAIALFASTMYLLFLPKILSILLWCKGTKEYGGFWRTLSLLLEVLF
VILAPVRLMHTYVVSFAFLCNEWVNSPORDDDSTSWGEAFKRGHCSOLLGLVWAG
MAMDURFLFWLAPITVLSLSPFVSIVSSRATVGLTKRWKLFLLIPEEISPPQVLYVD
TDRFLENRQRSLDDGFMAHVFNPSEFALATAMATARHASKRVLEIARDHRVQALNE
TPBKLNRDRLLVLLSDPTVWRLHFRVWNSPERYSYSSWVSYIEGILKLNPLALRPDAAS
Q"
7832..8059
/gene="ECS1428"
7832..8059
/gene="ECS1428"
/note="identical to YCEK_ECOLI gi|1787288 (Conserved in
E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="hypothetical protein"
/protein_id="BAB34851.1"
/db_xref="GI:13360892"
/translacion="MRLIVYSIMVTLSSGCSIIISRTIPQGGHGNQYYPGVQMDVRDS
AWRYVTILDLPFSLVFTLLPLDIDHHGPIE"
complement(8060..8437)
/gene="ECS1429"
complement(8060..8437)
/gene="ECS1429"
/note="similar to MSYB_ECOLI gi|1787289 percent identity
99 in 125 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="acidic protein MsyB"
/protein_id="BAB34852.1"
/db_xref="GI:13360893"
/translacion="MMTMYATLEAIDAAREEFLADNPIDAEDANVQQPNAQYVILQ
DGDIMQVFEFFADEGECEGLPMLSGEAGQSVFDGDDYDEIETIRQEWEOENTLHEWDEG

```

## Alignment Scores:

Pred. No.:	124	Length:	222605
Score:	119.50	Matches:	45
Percent Similarity:	44.59%	Conservative:	21
Best Local Similarity:	30.41%	Mismatches:	69
Query Match:	15.38%	Indels:	13
DB:	1	Gaps:	4

US-09-847-539a-6 (1-159) x AP002555 (1-222605)

Qy 15 GlyThrLeuThrAsnLeuLeuGlyAsnAlaProGlu-----LysLeuAlaLeu 30

|||||||:::|||||||

Db 184876 GGTACGCTGAATGATTTCTTGGTGCCATGACTGAGGATGATGTCCGTCGGAGGCACTG 184935

|||||||:::|||||||

Qy 31 ArgAsnGluGluArgAlaIleAspGluLeuLysGlnAlaIleGluAspLysGluAla 50

|||||||:::|||||||

Db 184936 CGCCGTTTGTAGCTGATGGTGAAGAGGTGGCGGTAAACGCGTCCGCGGTGGCACAGAAC 184995

/function="orf; Other or unknown (Phase or Prophage Related)"  
/note="Residues 1 to 191 of 193 are 94.76 pct identical to residues 1 to 191 of 191 from GenPept 118 : gi|2367120|gb|AAC74619.1| (AE000252) orf, hypothetical protein [Escherichia coli]"  
/codon\_start=1  
/transl\_table=11  
/product="putative tail fiber protein of prophage CP-933X"  
/protein\_id="AAG56009.1"  
/db\_xref="GI:12514849"  
/translation="MAFRMSEQPRTKIYNLLTGTNEFEGDVAIPPHGTLPANSTY IAPPDPAFVAVFNSDESHLVEDHKGTVYDVASGDALFISELGSIPENVTWLSPEGEFQKNGTAWKDAEAEKLFRIREABETKNSLMQVASEHIAPLQDAVDLEIATEEE ASLLCAWKRYRLLNRVDITSTADPWPVPEV"  
gene complement(7941. .8831)  
/gene="21921"  
CDS complement(7941. .8831)  
/gene="21921"  
/function="orf; Other or unknown (Phase or Prophage Related)"  
/note="Residues 1 to 289 of 296 are 84.42 pct identical to residues 1 to 289 of 292 from GenPept 118 : gi|7162108|emb|CAB76676.1| (Y19180) putative catalase [Salmonella typhimurium]"  
/codon\_start=1  
/transl\_table=11  
/product="unknown protein encoded by prophage CP-933X"  
/protein\_id="AAG56010.1"

Alignment Scores:  
Pred. No.: 5,34 Length: 10432  
Score: 119.50 Matches: 45  
Percent Similarity: 44.59% Conservative: 21  
Best Local Similarity: 30.41% Mismatches: 69  
Query Match: 15.38% Indels: 13  
DB: 1 Gaps: 4  
  
US-09-847-539a-6 (1-159) x AE005333 (1-10432)  
  
QY 15 GlyThrLeuThrAsnLeuLeuGlyAsnAlaProGlu-----LysLeuAlaLeu 30  
Db 4574 GCTACGCTGAATGATTTCTTGTCGCCACTGAGGATGATCGTCGGAGGCACCTG 4633  
  
QY 31 ArgAsnGluGluArgAlaIleAspGluLeuLysLysGlnAlaIleGluAspLysGluAla 50  
Db 4634 CGCCGTTTGGCTGATGTGTGAAGAAGTGGCGGTAAACCGCTCCCGTGGCCACAGAAC 4693  
  
QY 51 ThrThrAlaIleGluAlaAlaSerSerAsp-----AlaLeuGluAlaLeuAla 66  
Db 4694 ACGGCAGCCGCGAAGTACGCCACGATGCCAGCACATCAGCCGCTGAGGCGGCAACC 4753  
  
QY 67 AspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValLysAlaAspAsnAlaAla 86  
Db 4754 CATGCGACTGATGCTCGGCACTCAGCACGCGCAGCCAGCAGCTCAGCCGACAGGCGCG 4813  
  
QY 87 SerAspAlaLeuGluAlaLeuAlaAspGln---ThrAspAlaLeuGlnSerGluGluAla 105  
Db 4814 TCGTCGCTCAGTCACGCTCTCCACGCGGAGGACCGCATCAACAAAGGCTACTGAAGCA 4873  
  
QY 106 GluValValGlnSerAspAsnAlaAlaSerAspAlaThrGluLysAlaAlaThrProIle 125  
Db 4874 -----TCAAAAGTGTGCGCTGCGAGATCTCTCAAAAGCGCGCGGTACC 4921  
  
QY 126 AlaLeuAspValLysThrLysAspThrLysProValLysLysGluAlaArgGln 145  
Db 4922 AGTGGCGGTGCGCGCAAAACGTCAGAAACGAATGCGGCACTGTCAACAACAAATCAGCGCGC 4981  
  
QY 146 AsnValAsnThrLeuProThrThr 153  
Db 4982 ACTTCTGCATCCACCGCCGACCG 5005  
  
RESULT 23

AP002555 222605 bp DNA linear BCT 07-MAR-2001  
LOCUS Escherichia coli O157:H7 DNA, complete genome, section 6/20.  
DEFINITION AP002555 BA000007  
ACCESSION AP002555.1 GI:13360886  
KEYWORDS  
SOURCE Escherichia coli O157:H7 (strain:O157:H7, sub\_strain:RMD 0509952)  
DNA.  
Escherichia coli O157:H7  
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
Escherichia.  
REFERENCE 1 (sites)  
AUTHORS Makino,K., Yokoyama,K., Kubota,Y., Yutsudo,C.H., Kimura,S., Kurokawa,K., Ishii,K., Hattori,M., Tatsuno,I., Abe,H., Iida,T., Yamamoto,K., Ohnishi,M., Hayashi,T., Yasunaga,T., Honda,T., Sasakawa,C. and Shinagawa,H.  
TITLE Complete nucleotide sequence of the prophage VT2-Sakai carrying the verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7 derived from the Sakai outbreak  
JOURNAL Genes Genet. Syst. 74 (5), 227-239 (1999)  
MEDLINE 20198780  
REFERENCE 2 (sites)  
AUTHORS Ohnishi,M., Murata,T., Nakayama,K., Kuhara,S., Hattori,M., Kurokawa,K., Yasunaga,T., Yokoyama,K., Makino,K., Shinagawa,H. and Hayashi,T.  
TITLE Comparative analysis of the whole set of rRNA operons between an enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an Escherichia coli K-12 strain MG1655  
JOURNAL Syst. Appl. Microbiol. 23 (3), 315-324 (2000)  
MEDLINE 20557356  
REFERENCE 3 (sites)  
AUTHORS Yokoyama,K., Makino,K., Kubota,Y., Watanabe,M., Kimura,S., Yutsudo,C.H., Kurokawa,K., Ishii,K., Hattori,M., Abe,H., Iida,T., Yamamoto,K., Hayashi,T., Yasunaga,T., Honda,T., Sasakawa,C. and Shinagawa,H.  
TITLE Complete nucleotide sequence of the prophage VT1-Sakai carrying the Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli O157:H7 strain derived from the Sakai outbreak  
JOURNAL Gene 258 (1-2), 127-139 (2000)  
MEDLINE 20564182  
REFERENCE 4 (sites)  
AUTHORS Hayashi,T., Makino,K., Ohnishi,M., Kurokawa,K., Ishii,K., Yokoyama,K., Han,C.-G., Ohtsubo,E., Nakayama,K., Murata,T., Tanaka,M., Tobe,T., Iida,T., Takami,H., Honda,T., Sasakawa,C., Ogasawara,N., Yasunaga,T., Kuhara,S., Shiba,T., Hattori,M. and Shinagawa,H.  
TITLE Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic comparison with a laboratory strain K-12  
JOURNAL DNA Res. 8 (1), 11-22 (2001)  
MEDLINE 21156231  
REFERENCE 5 (bases 1 to 222605)  
AUTHORS Ohnishi,M., Kurokawa,K., Makino,K., Yasunaga,T., Shinagawa,H. and Hayashi,T.  
TITLE Direct Submission  
JOURNAL Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome Information Research Center; 3-1, Yamadaoka, Suita, Osaka 565-0871, Japan [E-mail:ken@gen-info.osaka-u.ac.jp, URL:http://www.gen-info.osaka-u.ac.jp/, Tel:81-6-6879-8365, Fax:81-6-6879-2047]  
COMMENT genome project.  
FEATURES Location/Qualifiers  
source 1..222605  
/organism="Escherichia coli O157:H7"  
/strain="O157:H7"  
/sub\_strain="RMD 0509952"  
/db\_xref="taxon:83334"  
74..607  
/gene="ECs1423"  
74..607  
/gene="ECs1423"  
/note="similar to B1045\_ECOLI gi|1787283 percent identity 100 in 177 aa (Conserved in E.coli K-12)"  
/codon\_start=1

Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamouasis,K.,  
 Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,  
 Welch,R.A. and Blattner,F.R.  
 Direct Submission  
 Submitted (22-Oct-2000) Laboratory of Genetics, University of  
 Wisconsin, 445 Henry Mall, Madison, WI 53706, USA  
 FEATURES  
 source  
 1. .10432  
 /organism="Escherichia coli O157:H7 EDL933"  
 /strain="EDL933"  
 /serotype="O157:H7"  
 /db\_xref="taxon:155864"  
 /note="enterohemorrhagic"  
 <1. .>10432  
 /note="O-island #52; Region of the EDL933 chromosome not  
 homologous to E. coli K-12 MG1655"  
 130. .1296  
 /gene="21915"  
 130. .1296  
 /gene="21915"  
 /function="putative structure; Structural component (Phase  
 or Propage Related)"  
 /note="Residues 8 to 366 of 388 are 81.05 pct identical to  
 residues 11 to 369 of 1132 from GenPept 118 :  
 gi|215125|gb|AA96553.1| (J02459) J (tail:host  
 specificity:1132) [bacteriophage lambda]"  
 /codon\_start=1  
 /transl\_table=11  
 /product="putative tail protein (partial) of prophage  
 CP-933X"  
 /protein\_id="AAG56004.1"  
 /db\_xref="GI:12514844"  
 /translation="MAARTREVEAKDNLKSQMMSVIDAIGEGPIEGPVKGLQSILV  
 NKTPTDNGNPIHGVTAVMRAGEQETPPCFESSGAETALGVETVAKRPTVTIT  
 SANIDRLVTFQVQLLETSKGRNHSVSLRLLOQRGNWVTEKDYTKNGTKTSQF  
 LASVILDLNPPFPIRMVRETTADSTTQQLNRLTLWSSYTEIIDVKQCPNTAIYGLQ  
 VDAREFGQQWTVYHIGRIITQVFNSTDPERTKTSGLWGLSKPAYSNNPAWCLWDM  
 LTHPRGMRKRGADVDKWLALYIAQYCDQTPDQPGFGTEPRMTFNAYLSQORKWD  
 VLSDPFSAMRCMPVWNGQTLTEVDQRPDSVWVPTPAVMMWMTAWGFATASAP"  
 1245. .3593  
 /gene="21916"  
 1245. .3593  
 /gene="21916"  
 /function="putative structure; Structural component (Phase  
 or Propage Related)"  
 /note="Residues 1 to 781 of 782 are 62.37 pct identical  
 to residues 376 to 1131 of 1132 from GenPept 118 :  
 gi|215125|gb|AA96553.1| (J02459) J (tail:host  
 specificity:1132) [bacteriophage lambda]"  
 /codon\_start=1  
 /transl\_table=11  
 /product="putative tail protein (partial) of prophage  
 CP-933X"  
 /protein\_id="AAG56005.1"  
 /db\_xref="GI:12514845"  
 /translation="MVVDNNGVFRYFSALKDRHTAVEVNTDPQNGWOTSTELVED  
 PPAILRYNKLKMGAFGTSRGOARHAGLWIKTLLLETQTVDFTLGSQLRHTPGD  
 IIEICNDNYAGTMGTGGRVLSIDAASRTLLDREVTLPTGATVNLINGSGKPVSDVI  
 TAPAPDRTOVSTLPDGVETGVWGLSLPCLRLFRVSVRENTDGTFAITAVQHPV  
 EKEAIVDNGAREPQSGNSLNVIPPAVOHLTVEVSADQYLQAQKWDTPRVKGVRF  
 SLRLTSGKTDARLVLTALTDATERHFSGLPIGEYTLTVRAINSYQOGEPAITTFRI  
 AAPAPSRITLPPGYQITATPLHAYDPTQVQFEFSEKRIADIRQVETAAARYLGS  
 LWXIASNIPKGDHYFYIRSVNTVSGKSAFVEAVGRASDDAEGYLDFFKGIKTHL  
 AOELMTQIDNGLOAPDLAEIRTSITNVSNEITQTVNKKLENSAIOQIOKVQVDNTN  
 NLSMWAYLQOMQDGRLYIAGIGAGIENTPAGMQSOVLLAADRILAMINPANGNTKPM  
 FYGQGDQIPMNEVFLKYLTAIPITSGGNPPAPSLTPDGLRTAKNADISGNVANSGLT  
 NVTINENCRVLGKLSAQOIGDLVTKGKAPPRDSRAPERPSETITVRYVDDQFPD  
 RQIVTAPAFSGAKHEKHEHTDIYSCRLLIVRKNAGIYINRTALDNTLIYSGVIDMPAG  
 HGHMTLEFSVSARLVNNNYPTASISDLLLVVVMKKATAGITIS"  
 3661. .4260  
 /gene="21917"  
 3661. .4260  
 /gene="21917"

/function="putative membrane; Other or unknown (Phage or  
 Propage Related)"  
 /note="Residues 1 to 199 of 199 are 75.87 pct identical to  
 residues 1 to 199 of 199 from GenPept 118 :  
 gi|17532789|gb|AA963231.1|AF151091\_2 (AF151091) Lom  
 [prophage P-EiBa]"  
 /codon\_start=1  
 /transl\_table=11  
 /product="putative outer membrane protein of prophage  
 CP-933X"  
 /protein\_id="AAG56006.1"  
 /db\_xref="GI:12514846"  
 /translation="MRKLCVILSAVVQVAAATPASAAEHQSTLSAGYLHASTNVPG  
 SDDLNGINVKRYEFMDALGILTSFYSADEQKTRYSDTRWHEDSYRNRFWSYMG  
 PSVRNWFSAYAMAGVAYSRTVSTGYDLRVTDNKGKTHDVLTCGDDGRHSNTSLAW  
 GAGVQFNPKESVAIDIAIEGSGCDWRDTDGFIVGVGYK"  
 4319. .7240  
 /gene="21918"  
 4319. .7240  
 /gene="21918"  
 /function="putative membrane; Not classified (Phage or  
 Propage Related)"  
 /note="Residues 1 to 498 of 973 are 95.78 pct identical to  
 residues 1 to 498 of 1122 from GenPept 118 :  
 gi|17876361|gb|AAC74454.1| (AE000234) putative membrane  
 protein [Escherichia coli]"  
 /codon\_start=1  
 /transl\_table=11  
 /product="putative membrane protein of prophage CP-933X"  
 /protein\_id="AAG56007.1"  
 /db\_xref="GI:12514847"  
 /translation="MNAVKISGVLDKGTGKPVENCITQLKARRNSATVNVNTVASEN  
 PDEAGRYGMDVEYGOYSVILLVEGPPSHAGTITVYEDSQPCLNDFLCAMTEDDVRP  
 EALREFELMEVEARNASAVQNTAAAKKSADASTSAREATHATDAADSAARASTS  
 AGQAASSQASSSSAGTASTKATEASKNAARAEKSKAAATSAAGAAKTSETNAAVSQ  
 SAATSEATATTKASASARDASAKSKAASSETSAASSASSASSASSASSASSAA  
 KTSETNAKSSSTAAEQSAASAKSKTAAALSASASTSAGASASTAAGSAESAAS  
 SASATTKAGATEQAASASASAAKTSETNAKSETSAESSKTAASASASAA  
 SSASAKDEATROASAAKSSATSTKATEAAGSATAAQAQSKSTAESATRAETAAKR  
 AEDIAVALEADASTTKGIIVQLSATNSTSELAATPKAKAAYLANGKYTAQDATN  
 TAQKGIIVQLSATNSTSELAATPKSKVAAYDLANGKYTAQDATN  
 SASSETLAATPKAKAANDNANGRPVSKVNGKALSSDITLTPKDIGLNTSTHFSFG  
 GAGWPKLATVTPMQASSVSITLIVAGNFGVSGPOQAGISELVLRAGNRPKITGALK  
 WORTSTGTFNFAWNTSGDTYDIYVAINATYATGVNIQWDTYSNASTVHTSPAYSANK  
 PEGITDGVYSLYTPSEQFYPPGAPIPWPSDTVPVSGYALMQCQTFDKTSAYPKLAAAYP  
 SGVIFDMRGWTIKGPKASGRAVLSQEDGIKSHTHSASASTDLGCTKTTSSFDYGTKS  
 TNNTCAHTHSVSGTAAASAGNTHSVTGASAVSQSGSVHKVYSAASVNTSAGAHT  
 HVSVGTAAASAGAHHTVIGIANTHVSALGSHGHTITVNAAGNAENTVNAIAFNIVRL  
 A"  
 complement(4822. .5571)  
 /gene="21919"  
 complement(4822. .5571)  
 /gene="21919"  
 /function="orf; Other or unknown (Phage or Propage  
 Related)"  
 /note="Residues 45 to 248 of 249 are 64.21 pct identical  
 to residues 5 to 205 of 206 from GenPept 118 :  
 gi|508991|gb|AA96556.1| (J02459) orf206 [bacteriophage  
 lambda]"  
 /codon\_start=1  
 /transl\_table=11  
 /product="unknown protein encoded by prophage CP-933X"  
 /protein\_id="AAG56008.1"  
 /db\_xref="GI:12514848"  
 /translation="MDAVVALLFAADACLVASSFEADADDADDAELADDAAAVFE  
 ALVSDAFVSDVFAAEALAAACSVAFVYVAVVVEADDAALSDAPPAVAAEAC  
 PALVDALDLNAAVFEPAEALCSAAVSEDLAFVSDVLAFAFELPAVAAEALLA  
 LEALVSDDFDAEFADSLAEASAFVYVAVVDAVDAVDAADCCDTAAAFVSDVFAAPALV  
 AAALFEDSAALAFDASVAFVDAVPALEAD"  
 7240. .7821  
 /gene="21920"  
 7240. .7821  
 /gene="21920"

gene

CDS

gene

CDS





```

QY 114 AlaSerAspAlaTrpGluLysAlaAlaThrProIleAlaLeuAspValLysLysThrLys 133
Db 2456 CCCGAGGAGAGTGAAGAGAGTCCAGGCCCTGTTGCACCTGAAGCTAAGAAAGCCCA 2515
QY 134 AspThrLysProValLysLysGluGluArgGlnAsnValAsnThrLeuProThrThr 153
Db 2516 GAGCAAAAGCAGCTCCCAAAAGAGAGCCCGCAAAAGGTGGAGTCA---CCCACGTCT 2572
QY 154 GlyGluGluSerAsnPro 159
Db 2573 GAGCCAGAGGATGAACCA 2590

RESULT 21
LOCUS AE006458/c
DEFINITION Escherichia coli O157:H7 EDL933 genome, contig 2 of 3, section 52
of 55.
ACCESSION AE006458 AE005174
VERSION AE006458.1 GI:13259569
KEYWORDS
SOURCE
ORGANISM
Escherichia coli O157:H7 EDL933.
Escherichia coli O157:H7 EDL933
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
1 (bases 1 to 9331)
Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
Grobeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamouisis,K.,
Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
Welch,R.A. and Blattner,F.R.
Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
Nature 409 (6819), 529-533 (2001)
21074935
PUBMED 11206551
REFERENCE 2 (bases 1 to 9331)
AUTHORS Plunkett,G. III.
TITLE Direct Submission
JOURNAL Submitted (12-FEB-2001) Laboratory of Genetics, University of
Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706-1577, USA
FEATURES
Location/Qualifiers
1..9331
/organism="Escherichia coli O157:H7 EDL933"
/strain="EDL933"
/serotype="O157:H7"
/db_xref="taxon:155864"
/note="enterohaemorrhagic"
complement(201..731)
/gene="z6025"
complement(201..731)
/function="orf; Other or unknown (Phage or Prophage
Related)."
/note="Residues 53 to 176 of 176 are 37.50 pct identical
to residues 13 to 140 of 140 from GenPept121:
dbj|BA36750.1 (AB016764) ORF4 [Escherichia coli]"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by cryptic prophage
CP-933P"
/protein_id="AAK16941.1"
/db_xref="GI:13259570"
/translation="MNVLRQAQVSSGRGEFTLGNVTSVIFNETDGRFLSSGSGGLL
TEFLYGNNGPEALRDLMLSDSGAQSQESIQDKISOCKFPVSSGNFQCPPEI
QCPTILPERPEGVFKNSDSSAVCCLEDFDAFSLASGSHPLREPIITASMIISPD
KCVDPKGNRTIKDS"
complement(888..1157)
/gene="z6026"
complement(888..1157)
/gene="z6026"
/function="orf; Other or unknown (Phage or Prophage
Related)."
/note="Residues 1 to 89 of 89 are 92.13 pct identical to

```

```

residues 41 to 129 of 129 from GenPept121:
gb|AAD25465.1|AF125520_60 (AF125520) hypothetical protein
[Bacteriophage 933W]"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by cryptic prophage
CP-933P"
/protein_id="AAK16942.1"
/db_xref="GI:13259571"
/translation="MNILKLMORLCGHGKHDDREDGELLTAQLRLGPADLLESDENG
IIPQDRVITQVILADKKQIQCVVRPLQILRADCTWENIGMK"
complement(1159..2472)
/gene="z6027"
complement(1159..2472)
/gene="z6027"
/function="putative structure; Structural component (Phage
or Prophage Related)."
/note="Residues 1 to 437 of 437 are 94.96 pct identical to
residues 1 to 437 of 437 from GenPept121: dbj|BAB19560.1|
(AP000400) tail fiber protein [Escherichia coli O157:H7]"
/codon_start=1
/transl_table=11
/product="putative tail fiber protein of cryptic prophage
CP-933P"
/protein_id="AAK16943.1"
/db_xref="GI:13259572"
/translation="MTVKISGLKDGTKPVQNCITVLKARRTSSTVVVNTVAASENPD
EAGYSDMDEVGOYSVILLVVEGPPSHAGTITVYEDSQPTLNDFLGMSDDVYRPEA
LRRELWVEEAARHAEEAKKNAGEAETSARNAGISAQAEANAADTSAGDASESAR
AKAESAESAQSAESAESAESAESAESAESAESAESAESAESAESAESAESAESAESA
TGPAGATGERGPAGDAGPAGQPGKGRGERGETGLTGNAGPQPGKDGKGERGD
PKGETGAAGPVGATPGQPGKGRGERGETGLTGNAGPQPGKDGKDGKGERGD
TFLAPKDTTRVQGFQHLQVRFGDPQWQVKGDLDEVGSDTGRTCE"
complement(2624..3223)
/gene="lomp"
complement(2624..3223)
/gene="lomp"
/function="putative membrane; Other or unknown (Phage or
Prophage Related)."
/note="Residues 1 to 199 of 199 are 96.48 pct identical to
residues 1 to 199 of 199 from GenPept121: dbj|BAB19561.1|
(AP000400) outer membrane protein Lom precursor
[Escherichia coli O157:H7]"
/codon_start=1
/transl_table=11
/product="putative Lom-like outer membrane protein of
cryptic prophage CP-933P"
/protein_id="AAK16944.1"
/db_xref="GI:13259573"
/translation="MRKLCVILSAVYVQVAAATPASAEHQSTLSAGYLHASTNVPG
SDDLGNLKYRYEEMDALGLITSFYANAEDOKTRYSDTRHEDSVRNRFWSVMAG
PSVRNWFSAAYAGVAYSRVSFFGYDLRVTONKTKTHDVLGSDSDGRHSNTSLAW
GAGVQFNPTESVTIDLAYEGSGSDWRSDAFIVGIGYRF"
complement(3291..5639)
/gene="z6029"
complement(3291..5639)
/gene="z6029"
/function="putative structure; Structural component (Phage
or Prophage Related)."
/note="Residues 1 to 782 of 782 are 99.87 pct identical to
residues 25 to 806 of 806 from GenPept121: dbj|BAB19562.1|
(AP000400) host specificity protein J [Escherichia coli
O157:H7]"
/codon_start=1
/transl_table=11
/product="putative tail component of cryptic prophage
CP-933P"
/protein_id="AAK16945.1"
/db_xref="GI:13259574"
/translation="MVDNDNGVGFYRFSALKDRHTAVEVNTDPQNGWQTSTELVED
PEATLRYGRNLLKMDAFGCTSRGQHRAGLWIKTGLLETQTVDTLGSQGLRHTPGD

```

```
/note="unknown"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34182.1"

Alignment Scores:
Pred. No.: 34 Length: 297816
Score: 128.50 Matches: 49
Percent Similarity: 38.59% Conservative: 22
Best Local Similarity: 26.63% Mismatches: 66
Query Match: 16.54% Indels: 47
DB: 1 Gaps: 5

US-09-847-539a-6 (1-159) x AP002553 (1-297816)
Qy 15 GlyThrLeuThrAsnLeuLeuGlyAsnAlaProGlu-----LysLeuAlaLeu 30
Db 83710 GGTACGCTGAATGATTTCTCGTGCCATGACGAGGATGATCCCTCGGAGGCACTG 83769
Qy 31 ArgAsnGluGluArgAlaIleAspGluLeuLysLysGlnAlaIleGluAspLysGlu--- 49
Db 83770 CGTGGTTTGAACAGATGTTGGAAGAAGCAGCGCTCAGCAGAGGAGCGCAAGAAGAT 83829
Qy 49 ----- 49
Db 83830 GCGGGAAGCAGAGAGCTCGGCGGAGAAATCGCGGCATATCAGCCAGTAAGCGGAAGCG 83889
Qy 50 -----AlaThrAlaIleGluAlaAlaSerSerAspAlaLeuGluAlaLeu 65
Db 83890 AGCCGCCAAATGCTGATATTCTACGAGAGGATGATCGGAGTCAGCCCGCAGCGCGCA 83949
Qy 66 AlaAspGlnThrAspAlaLeuGlnSerGluGluAlaValValLysAlaAspAsnAla 85
Db 83950 GAAAGTCAGCCTCTGAAAGAGAGTCAGAGAGCGCTCTGCTCAGCCTCTGAGGCC 84009
Qy 86 AlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAla 105
Db 84010 GCTCAAAAGCCAGTGTAGTATTACAAAGTCACACAGATCCGAGTTGTCAAAAAGACG 84069
Qy 106 GluValValGlnSerAspAsnAlaAlaSerAspAlaTrpGluLysAlaAlaThrProIle 125
Db 84070 GCAGAAAGTCAGCCGGTATATGACGCCAGGATGCA----- 84105
Qy 126 AlaLeuAspValLysLysThrLysAspThrLysProValValLysLysGluGluArgGln 145
Db 84106 AGCACTCAACAGAAAAGCCCGGGAATCAGAGAAAGCCACAGTCAGCGGAACAAAGC 84165
Qy 146 Asn-----ValAsnThrLeuProThrThr-----GlyGlu 155
Db 84166 AGAATACGGCGGAAGACGCCGTAAACAGAAATTCCACCGTGTGGGGCTCCCGGACCA 84225
Qy 156 GluSerAsnPro 159
Db 84226 AAGGGGAACCG 84237

RESULT 20
PMU19361
LOCUS PMU19361 3763 bp mRNA linear VRT 28-OCT-1995
DEFINITION Petromyzon marinus neurofilament subunit NF-180 mRNA, complete cds.
ACCESSION U19361
VERSION U19361.1 GI:632548
KEYWORDS sea lamprey.
SOURCE Petromyzon marinus
ORGANISM Petromyzon marinus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Hyperoartia;
Petromyzontiformes; Petromyzontidae; Petromyzon.
1 (bases 1 to 3763)
JACOBS,A.J., Kamholz,J. and Seizer,M.E.
The single lamprey neurofilament subunit (NF-180) lacks
multiphosphorylation repeats and is expressed selectively in
projection neurons

JOURNAL Brain Res. Mol. Brain Res. 29 (1), 43-52 (1995)
MEDLINE 95287814
REFERENCE 2 (bases 1 to 3763)
AUTHORS Jacobs,A.J.
TITLE Direct Submission
JOURNAL Submitted (03-JAN-1995) Alan J. Jacobs, Department of Neuroscience,
University of Pennsylvania, 452 Stemmler Hall, 3600 Hamilton Walk,
Philadelphia, PA 19104-6087, USA
FEATURES
Location/Qualifiers
1..3763
/organism="Petromyzon marinus"
/db_xref="taxon:7757"
/clone="clones LIF5, LIF13, and LIF22"
/tissue_type="central nervous system"
/clone_lib="Lambda gt11 cDNA library from larval lamprey
CNS"
/dev_stage="larva"
122..3454
/note="180 kDa neurofilament subunit"
/codon_start=1
/evidence=experimental
/product="NF-180"
/protein_id="AAA80106.1"
/db_xref="GI:632549"
/translacion="MSYTESVTCYRRAYADASRGSKRLLSLTPSSGFRSQCWRCGS
SGGSGSLVSLQRRPNYASLSHGYSTPALNSAADSLDPHSQSSVSFSGDFKHSRVNEKQ
LQDLNDRFAGYIEKHVDLEQKNKELETEISAYRKOSQSPARGGGVSDVTEQEIKEIR
VDITDDINGEKTQVQIEHLDREIQREKTDDERLNRNTEALINAFKRNVDITSLV
RMDKRTQSLDEITFLKKNHEEVDLELLAQTSSTVSVERKDFAVPEITAAALRETR
GOLGOSARNIETAEWFKGFSQLEAAEONNDIATRSAAKEETEHRRKLQMRCTELD
ALAGTESLEROLSEMERHOSQDNLQDAQOLENLRNTKWMARHLHREYODLLNV
KHALDTEIARYKLLOGEIRYSSGRLPTPAKPPKAPSAKPAKAKVEKKVYKKEPEK
VESEPISQDLTDLEIRYVMEAKAPVVSAAEKDEEEEEEEEEEEEEEEEEEEEEEE
EDGRKEGEAAEEEEEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEE
GEAEAEVEEVEEAEIKBAEAAEAEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEE
AEVEEVEEAEAEAEAEAEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEE
EDVEAESEEEEEEDSKEADEAEDEAEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEE
AEBAKEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAE
EPAKAPKPAKPAKVEPTSEPEDEPKAEVVEKKGABAPKPAKPAKAAKPAKPAK
EPEPEPTSEPEKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAK
EKDDDEEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEE
DKAEPKAEKHSPEERKPIKEIAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAK
ETDQSTQDESVEEIEETIVSTKKVEKVAAPSKPEKPAKPAKPAKPAKPAKPAK
IETVSTTKKHEKRAPVDSK"
BASE COUNT 1177 a 918 c 1248 g 420 t
ORIGIN
Alignment Scores:
Pred. No.: 11 Length: 3763
Score: 122.50 Matches: 37
Percent Similarity: 51.59% Conservative: 28
Best Local Similarity: 29.37% Mismatches: 50
Query Match: 15.77% Indels: 11
DB: 5 Gaps: 5

US-09-847-539a-6 (1-159) x PMU19361 (1-3763)
Qy 38 AspGluLeuLysLysGlnAlaIleGluAspLysGluAlaThrThrAlaIleGluAla 57
Db 2234 GACCAACAGAGAGAGAGGAGGTTAAGAGAGAGAGGTCACC----- 2275
Qy 58 SerSerAspAlaLeuGluAlaLeuAlaAsp---GlnThrAspAlaLeuGlnSerGluGlu 76
Db 2276 AAGTCAGATGCTGAAGAGAGCTGAGGCTGAAGCAGAGAGGCTGCAAGTCAGAGAG 2335
Qy 77 AlaAlaValValLysAlaAspAsnAlaAla-----SerAspAlaLeuGluAlaLeuAla 94
Db 2336 GAAGCGCTGAAGAGAGCAAGATGAAGCTGAGGAGGAGAGAGGAGGAGGCTGTT 2395
Qy 95 AspGlnThrAspAlaLeuGlnSerGlu---GluAlaGluValValGlnSerAspAsnAla 113
Db 2396 GAAGAAACTGAAGCAGCACTGAAGAGAGCTGAAGCAGAGGAGGCTCAGACGATGACAAA 2455
```



```

/db_xref="GI:12513757"
/translation="WLSPYŠVNLGCSWNSLTRLNLSFDPNRVLSSVRDAAVHSDNGAQQV
KGNIRYVATDNRFCVYKRESHSGCTFNLHLRLGMPKGEISRKIEVMLNAPSPVSAAM
ERGIHVHNSRPDPDPPDYAPPELPSVDYNRLSPGVNGKGNVAYVEDADATKVLKM
FTTOSNEEVTVRCFNQYAGSAEKIYGNNGDIIIGIRMDKINGESLNISSLPAPQ
AHAIYDMFDRLEQKILFVDTTETNVLYDRAKNFNPDISSVNVSDRSWSESQIMQ
SVHGGKODLISVLSKI"
gene
6462..7160
/genes="z0990"
6462..7160
/genes="z0990"
/function="orf; Other or unknown (Phage or Prophage
Related)"
/notes="No significant matches"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AAG55143.1"
/db_xref="GI:12513758"
/translation="WRPTSLNLVHLQSSRSSMSDITDSLKASSVQWIKNPQLRP
QQTIDINIQOIEAALDKIGSTETGRVLLNAIESIRLKSSETVVIHLNLSRLGMARHD
IDAEHRGTGSDPHCNLNAVEYPCGEGISVDFHATIVFHELLHFHNLNGLRKVES
SRAESQKSPLLLEARTVGLGFSEVLSENKFHEIGMPRTSYPPXDSALIHDDNT
VSLGPGQVRLHPLL"
gene
complement(7668..8144)
/genes="vphB"
/notes="z0992"
complement(7668..8144)
/genes="vphB"
/function="orf; Unknown function"
/notes="Residues 1 to 158 of 158 are 99.36 pct identical to
residues 1 to 158 of 158 from Escherichia coli K-12 Strain
MG1655: B0773"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAG55144.1"
/db_xref="GI:12513759"
/translation="WKLISNLDRLDGLKPLHRVFNMGYDGNISPHLAWDDVPAGTK
SFVVCYDPDATGSGMHWVNLNPLADTRVLPPQGGSLVAMPDGVLTQTRVDFGKTG
YDGAAPPKGETHRYIFTVHALDVRIDVDEGASGAVGNFVHFHSLASASITAMFS"
gene
complement(8203..9492)
/genes="bioA"
/notes="z0993"
complement(8203..9492)

Alignment Scores:
Pred. No.: 1.41 Length: 13501
Score: 128.50 Matches: 49
Percent Similarity: 38.59% Conservative: 22
Best Local Similarity: 26.63% Mismatches: 66
Query Match: 16.54% Indels: 47
DB: 1 Gaps: 5

US-09-847-539A-6 (1-159) x AE005258 (1-13501)
QY 15 GlyThrLeuThrAsnLeuLeuGlyAsnAlaProGlu-----LysLeuAlaLeu 30
Db 977 GGTACCGCTGAATGATTTCTCGTGCCATGCAGGAGGATGTCGCCGTCGGAGGCACGTG 1036
QY 31 ArgAsnGluGluArgAlaIleAspGluLeuLysLysGlnAlaIleGluAspLysGlu--- 49
Db 1037 CGTCGCTTTGAACAGATGTTGGNAGAGCAGCCGCTCAGCAGAGGAGCGAAGAGAAT 1096
QY 49 ----- 49
Db 1097 GCCGGAGAACGACAGACGTCCCGGAGGAATGCCGGCATATCAGCCAGTAAGCGGAAGCG 1156
QY 50 -----AlaThrThrAlaIleGluAlaAlaSerSerAspAlaLeuGluAlaLeu 65
Db 1157 AGCGCCGCAAAATGCTGATCTACTCAGCAGAGGATGCATCGGATCGACGCGGCGCGCA 1216
QY 66 AlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLysAlaAspAsnAla 85

```

```

Db 1217 GAAAGTGCAGCCCTCTGCAAGAAGTGCAGGAAGCGCTCTGCTCCTCAGGCTCTGAGGCC 1276
QY 86 AlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAla 105
Db 1277 GCTCAAAAAGCCAGTGTAGTCTATTACAAGTGCACAGATGCCGAGTTGTCTCAAAAAGACG 1336
QY 106 GluValValGlnSerAspAsnAlaAlaSerAspAlaTrpGluLysAlaAlaThrProIle 125
Db 1337 GCAGAAAGTGCAGCCGCTAATGCAGCAGGATGCA----- 1372
QY 126 AlaLeuAspValLysLysThrLysAspThrLysProValValLysLysGluGluArgGln 145
Db 1373 ACGACCTCAACAGAAAAAGCCGGGAATCAGCAGAAAGCGCACAGTCAGCGGAACAAAGC 1432
QY 146 Asn-----ValAsnThrLeuProThrThr-----GlyGlu 155
Db 1433 AGAATAGCGCGGAGAGCGCGTAAACAGAAATTCACCCGCTGGTGGGCGCTCCCGGACCA 1492
QY 156 GluSerAsnPro 159
Db 1493 AAGGGGGAACCG 1504
RESULT 19
LOCUS AP002553 297816 bp DNA linear BCT 07-MAR-2001
DEFINITION Escherichia coli O157:H7 DNA, complete genome, section 4/20.
ACCESSION AP002553 BA000007
VERSION AP002553.1 GI:13360211
KEYWORDS
SOURCE Escherichia coli O157:H7 (strain:O157:H7, sub_strain:RIMD 0509952)
ORGANISM Escherichia coli O157:H7
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE 1 (sites)
AUTHORS Makino,K., Yokoyama,K., Kubota,Y., Yutsudo,C.H., Kimura,S.,
Kurokawa,K., Ishii,K., Hattori,M., Tatsuno,I., Abe,H., Iida,T.,
Yamamoto,K., Ohnishi,M., Hayashi,T., Yasunaga,T., Honda,T.,
Sasakawa,C., and Shinagawa,H.
Complete nucleotide sequence of the prophage VT2-Sakai carrying the
verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7
derived from the Sakai outbreak
Genes Genet. Syst. 74 (5), 227-239 (1999)
20198780
REFERENCE 2 (sites)
AUTHORS Ohnishi,M., Murata,T., Nakayama,K., Kuhara,S., Hattori,M.,
Kurokawa,K., Yasunaga,T., Yokoyama,K., Makino,K., Shinagawa,H. and
Hayashi,T.
Comparative analysis of the whole set of rRNA operons between an
enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an
Escherichia coli K-12 strain MG1655
Syst. Appl. Microbiol. 23 (3), 315-324 (2000)
20557356
REFERENCE 3 (sites)
AUTHORS Yokoyama,K., Makino,K., Kubota,Y., Watanabe,M., Kimura,S.,
Yutsudo,C.H., Kurokawa,K., Ishii,K., Hattori,M., Abe,H., Iida,T.,
Yamamoto,K., Hayashi,T., Yasunaga,T., Honda,T., Sasakawa,C. and
Shinagawa,H.
Complete nucleotide sequence of the prophage VT1-Sakai carrying the
Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli
O157:H7 strain derived from the Sakai outbreak
Gene 258 (1-2), 127-139 (2000)
20564182
REFERENCE 4 (sites)
AUTHORS Hayashi,T., Makino,K., Ohnishi,M., Kurokawa,K., Ishii,K.,
Yokoyama,K., Han,C.-G., Ohtsubo,E., Nakayama,K., Murata,T.,
Tanaka,M., Tobe,T., Iida,T., Takami,H., Honda,T., Sasakawa,C.,
Ogasawara,N., Yasunaga,T., Kuhara,S., Shiba,T., Hattori,M. and
Shinagawa,H.
Complete genome sequence of enterohemorrhagic Escherichia coli
O157:H7 and genomic comparison with a laboratory strain K-12
DNA Res. 8 (1), 11-22 (2001)

```

```
Escherichia.
1 (bases 1 to 13501)
Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamouis,K.,
Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
Weich,R.A. and Blattner,F.R.
Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
Nature 409 (6819), 529-533 (2001)
21074935
PUBMED
11206551
2 (bases 1 to 13501)
Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamouis,K.,
Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
Weich,R.A. and Blattner,F.R.
Direct Submission
Submitted (22-OCT-2000) Laboratory of Genetics, University of
Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
FEATURES
Location/Qualifiers
1..13501
/organism="Escherichia coli O157:H7 EDL933"
/strain="EDL933"
/serotype="O157:H7"
/db_xref="taxon:155864"
/note="enterohaemorrhagic"
misc_feature
<1..7576
/note="O-island #36; Region of the EDL933 chromosome not
homologous to E. coli K-12 MG1655; Cryptic prophage
CP-933k; includes one copy of the 13 bp direct repeat that
flanks the prophage"
66..665
/gene="lomk"
/note="20981"
66..665
/gene="lomk"
/function="putative membrane; Other or unknown (Phage or
Prophage Related)"
/note="Residues 1 to 199 of 199 are 72.86 pct identical to
residues 1 to 199 of 199 from Genpept 118 :
gi|7532789|gb|AAFG6231.1|AF151091_2 (AF151091) lom
[prophage P-Eiba]"
/codon_start=1
/transl_table=1
/product="putative outer membrane protein of prophage
CP-933k"
/protein_id="AAG55137.1"
/db_xref="GI:12513752"
/translation="MRKVCAAILSAICLAVSGVPWASEHQSTLSAGYLHASTDAPG
SDDLINGINVKRYEFTDPLGLITSPSYANAEDKQTHYSDTRWHEDEVRFVWFVSMAG
PSVRVNEFSAYAMAGVAYSRVTSFGDYFRVTKNKRKTHDVLIGSDDARYSNTSLAW
GAGVQNPTESVAVDVAVEXSGSDWRDGTGFIQVGVGYKF"
719..2041
/gene="20982"
719..2041
/gene="20982"
/function="putative structure; Structural component (Phage
or Prophage Related)"
/note="Residues 164 to 440 of 440 are 68.79 pct identical
to residues 381 to 645 of 645 from Genpept 118 :
gi|4585436|gb|AAD25464.1|AF125520.59 (AF125520) putative
tail fiber protein [Bacteriophage 933w]"
/codon_start=1
/transl_table=1
/product="putative tail component of prophage CP-933k"
/protein_id="AAG55138.1"
/db_xref="GI:12513753"
/translation="MGMAAVQISGLVKDGACKPIONCTIQLKARNSTXYVVNTVASE
NPDEGRYSMDVEVQYSVTLVVGFPFPHSHAGTISVYEDSQPGLNDFLGAMTEDDAR
PEALRRFQMVEEAARHAEAKNAGEAETSARNAGTSASAKAEASANAADTSAEDASE
SARQAEASAAKKSSEASSSSASEAOKASELSQATDAELSKKTSKTSAGNAARDA
TTSTEKARSAESQASQSRIAAEDAVNRITPVYGGPKGEPGAGPGKDGKE
RGDTGPAGATGEPGGDTGPAGQPGKGDGGERGETGLTGNAGPGKGTGAAGPA
GPGPKGTGAAGPVGATGPGQPGKGTGCTQIRPLGPMRIETNSYGFPGTDCALI
TGLTFLDPKDATQVQMFQHLQVRFQDGPWQDVKGLDEVGSDTGRTE"
2004..2312
/gene="20984"
2004..2312
/gene="20984"
/function="orf; Other or unknown (Phage or Prophage
Related)"
/note="Residues 1 to 102 of 102 are 98.03 pct identical to
residues 28 to 129 of 129 from Genpept 118 :
gi|4585437|gb|AAD25465.1|AF125520.60 (AF125520)
hypothetical protein [Bacteriophage 933W]"
/codon_start=1
/transl_table=1
/product="unknown protein encoded by prophage CP-933k"
/protein_id="AAG55139.1"
/db_xref="GI:12513754"
/translation="MKWAVIOAELENDMNLRLKMQSLCGCKHDCGNGRSLTAQLR
LGPADILESDENGIIPQDRVITQVWILDADKKQTCVVRPLQILRADGTWENIGMK
"
2489..3469
/gene="20985"
2489..3469
/gene="20985"
/function="orf; Other or unknown (Phage or Prophage
Related)"
/note="Residues 13 to 326 of 326 are 55.73 pct identical
to residues 18 to 331 of 336 from Genpept 118 :
gi|6960367|gb|IAAF33527.1| (AF170176) hypothetical protein
predicted by Glimmer [Salmonella typhimurium LT2]"
/codon_start=1
/transl_table=1
/product="unknown protein encoded by prophage CP-933k"
/protein_id="AAG55140.1"
/db_xref="GI:12513755"
/translation="MLSPRTTFHNSVNIQSSPCQTVSFAGKEVLKVIDEKTPILF
QWEPNPERYKKDEVPIVNTKQHPYLDNVTNAARIESDRMIGIFVDGFSVNOKTAFS
KLEDFPENVMIIYREDVDVDSMYDRKSLDIYHDIICEQRLTEDRDEYLLNLEKELR
EISRAQSLISMYAKKRNHAWDFFRNLALAKAGEIFRCTYTNKHNLGSPFGCIIYLD
MDMLTKLGTIYAPDGMISHVDRNDSVNIENSALIVNRSNHPALLGSLGSLFMHKSVD
AHYYDGLGKGVKYYFNFTPLHNYHFCDTIEFNHPNIMNTSQTCSW"
3530..4522
/gene="20986"
3530..4522
/gene="20986"
/function="orf; Other or unknown (Phage or Prophage
Related)"
/note="No significant matches"
/codon_start=1
/transl_table=1
/product="unknown protein encoded by prophage CP-933k"
/protein_id="AAG55141.1"
/db_xref="GI:12513756"
/translation="WKIPSLQSNFNSAPAGYSAPIAPNRAENAYADYVLDIGKRPL
SAADLSNVYSVIRAVDRSRRLDQHTVDMIGNTVLDALRSOTFRDQVSYGTHNEK
VHIGCIKYREYELNEBSSVKIDDIQSLTCNELYEDVGOEPIFPICEAGENDNEEY
VSFVADPTDSYEMPSQOGLIHEIITHVTGSSDPGSDSNIELGPTTEILARRVAQELG
WSVPDFPGYAEAREAHRLRLNALRQAAHRENERAFPERLGTISDRYEASPDFT
EYSAVSNIGYGIQOHDHDPGLAINDNLQDANQIQLYHGAPYIFTFGDVQDKHNOQ"
5350..6231
/gene="20989"
5350..6231
/gene="20989"
/function="orf; Other or unknown (Phage or Prophage
Related)"
/note="No significant matches"
/codon_start=1
/transl_table=1
/product="unknown protein encoded by prophage CP-933k"
/protein_id="AAG55142.1"
```

<b>COMMENT</b>					
Twenty degenerate tandem repeats of twentyfour base pairs each start at position 953 and end at 1432.					
<b>FEATURES</b>					
source	1..3838	/organism="Plasmodium falciparum"			
		/db_xref="taxon:5833"			
sig_peptide	665..733				
CDS	665..1594	/note="S antigen signal peptide"			
		/note-"S antigen precursor"			
		/codon_start=1			
		/protein_id="AAA29758.1"			
		/db_xref="GI:160671"			
mat_peptide	734..1591	/translation="NNRIISVGFYLFLLYLIIYKYTKVGKNTDQEISNIYGTN GFLNGKNGKNXYEDLOEREGENDEEHSESDNDDENEIIVQDEAPKSI LKSDEAEALKSDEAEALKSDEAEALKSDEAEARKSDEAEALKSDEAEALKSDEA SDEAEALKSDEAEALKSDEAEALKSDEAEARKSDEAEALKSDEAEARKSDEAE EAERAKSDEAEALKSDEAEALKSDEAEARKSDEAEARKSDEAEARKSDEAEAR EGPKGTGGPGSGGSHHNKKSKKSTMMNLILM"			
		/product="S antigen"			
BASE COUNT	1601 a 270 c 504 g 1463 t				
ORIGIN	1 bp upstream of EcoR1 site.				
Alignment Scores:					
Pred. No.:	0-228	Length:	3838		
Score:	131.50	Matches:	43		
Percent Similarity:	57.14%	Conservative:	29		
Best Local Similarity:	34.13%	Mismatches:	45		
Query Match:	16.92%	Indels:	9		
DB:	3	Gaps:	6		
US-09-847-539A-6 (1-159) x PFASAT7 (1-3838)					
Qy	26	GLuLYsLeuAlaLeuArgAsnGluArGAlaIleAlaSpGlueLuLeuLYsGlnAlaIle	45		
Dd	959	GAGGCAGAGCCACTAAAAAGTGATGAG-----GCAGGCCACTAAAAAGTGATGAGGCA	101		
Qy	46	GLuAspLYsGLuAlaThrThrAlaIleGluAlaAlaSerSerAspAlaLeuGluAla---	64		
Dd	1013	GAGGCACGAAAAAGTGTATGAGGCA---GAGGCCACTAAAAAGTGATGAGCGACGACCA	1067		
Qy	65	LeuAlaAlaSpGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLYsAlaAspAsn	84		
Dd	1070	AAAAGTGATGAGCGCAGGCGCACTAAAAAGTGATGAGCGCAGGCCACTAAAAAGTGATGAG	112		
Qy	85	Ala-----AlaSerAspAlaLeuGluAlaLeu---AlaAspGlnThrAspAlaLeu	100		
Dd	1130	GCAGAGGCACGAAAAAGTGATGAGGCGAGGCGCACTAAAAAGTGATGAGCGCAGGCGCAC	118		
Qy	101	GlnSerGluGluAlaGluAlaValGlnSerAspAsnAlaAlaSerAspAlaTrpGluLYs	120		
Dd	1190	AAAAGTGATGAGCGCAGGCGCACTAAAAAGTGATGAGCGCAGGCGCGCAAAAAAGTGATGAG	124		
Qy	121	AlaAlaThrProIleAlaLeuAspValLYsLYsThrLYs---AspThrLYsProValVal	139		
Dd	1250	GCAGAGCCACCAGAAAGTGATGAGGCGAGGCGCAGAACGAAAGTGATGAGCGCAGGCGCA	1301		
Qy	140	LYsLYsGluGluArgGln	145		
Dd	1310	AAAAGTGATGAGCGCAGAG	1327		
RESULT 18					
AE005258					
LOCUS					
Escherichia coli O157:H7 EDL933 genome, contig 1 of 3, section of 155.					
DEFINITION					
ACCESSION AE005258 AE005174					
VERSION					
KEYWORDS					
SOURCE					
ORGANISM					
Escherichia coli O157:H7 EDL933.					
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae					

```
US-09-847-539A-6 (1-159) x SSCMLB1 (1-1380)
QY 19 AsnLeuLeuGlyAsnAla-----ProGluLysLeuAlaLeu 30
Db 601 AATCTTTGGGAATGCTAAAGATCAATTAGATAAATCTTACTACAGAAAAGAAAGATTG 660
QY 31 ArgAsnGluGluArgAlaIleAspGluLeuLysLys----- 42
Db 661 TCTAAGATAAAGAGCAATAGATGAAGAAAATAAGATCTTGGAGCAAGTGTACCCGT 720
QY 43 -----GlnAlaIleGluAspLysGluAlaThrThrAlaIleGluAla 57
Db 721 ACAAAACCGTGATTAGAGCGAGCTCGTGACGCTAAGAAAGCCACAGCAAGATAGCT 780
QY 58 SerSerAspAla-----LeuGluAlaLeuAla 66
Db 781 GAAACACAGCTAAAGTTGACAAAGCTTGNAGAGAAAACAAGTCTTGGAGCCAGCCGT 840
QY 67 AspGlnThr-----AspAlaLeuGlnSerGluGluAla 78
Db 841 AAACGTACAAACCGTGATTAGAGCGAGCTCGTGACGCTAAGAAAGCCACAGACAGAA 900
QY 79 ValValLysAlaAsp----- 83
Db 901 TTGGCTAAAGCTTAATGAGCTAAACCAAAATCTTAGAAGCAAGCCGACAGCAACCGT 960
QY 84 -----AsnAlaAlaSerAspAlaLeuGluAla-----LeuAla 94
Db 961 GATTAGAGCGAGCTCGTGACCGCTAAGAAAGCAGTTGATGACAGAACTTGCTAAACTTAA 1020
QY 95 AspGlnThrAspAlaLeuGlnSerGluGluAlaGluValValGlnSerAspAlaAla 114
Db 1021 GCTGAGCTGAGCGCTTAAGAGCAATTTGGCTAAACAAGCTCAAGAAATTTGAAAACCT 1080
QY 115 SerAspAlaTrpGluLysAla----- 121
Db 1081 AAGAATCAAAAGAAAAGACCCAGCAAGCAACTCAAACTCCAGAAAACCCAGAGTACCT 1140
QY 122 -----AlaThrProIleAlaLeuAspVal 129
Db 1141 GCTAAACCAAGCATGCTTGGACAGGATTAACACCTGTACACCAATCGCTAAGACAGA 1200
QY 130 LysLysThrLysAspThrLysProVal-----ValLys 140
Db 1201 AAGAAAACCT--GATGTTAAAGCTGCGCTAAAGCAAAACATGGTGCTACTGACGTTAG 1257
QY 141 LysGluGluArgGlnAsnValAsnThrLeuProThrThrGlyGluGluSerAsnPro 159
Db 1258 AAGATGAGAAGAAA-----CTTCATCAACTGGCGAAACTGTGAACCCA 1302
RESULT 15
LOCUS I09107 798 bp linear PAT 02-DEC-1994
DEFINITION Sequence 4 from Patent WO 8810306.
ACCESSION I09107
VERSION I09107.1 GI:588179
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE Unclassified.
AUTHORS 1 (bases 1 to 798)
TITLE CLONED PROTEIN G VARIANT GENES AND THE PROTEIN G VARIANTS EXPRESSED THEREFROM
JOURNAL Patent: WO 8810306-A 4 29-DEC-1988;
FEATURES Location/Qualifiers
source 1..798
BASE COUNT 277 a 161 c 177 g 183 t
ORIGIN
Alignment Scores:
```

```
Pred. NO.: 0.012 Length: 798
Score: 139.00 Matches: 58
Percent Similarity: 35.65% Conservative: 24
Best Local Similarity: 25.22% Mismatches: 64
Query Match: 17.89% Indels: 84
Db: 6 Gaps: 8
US-09-847-539A-6 (1-159) x I09107 (1-798)
QY 4 ProIleGluGlnProArgIleIleProAsnGlyThrLeuThrAsnLeuLeuGlyAsn 23
Db 100 CCAATCGAAGATACCCCAATTATTCTGTAATGGTGTGTAATTAACCTTAATCTTCTGGGAAT 159
QY 24 AlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaIleAspGlu----- 39
Db 160 TCAGAGACAACACTGGCTTGGTAATGAAGAGAGTGTCTACAGCTGGGTACCCACTTACT 219
QY 40 -----LeuLysLysGlnAlaIleGluAspLysGluAlaThr 51
Db 220 AAGACTGACACTTACAAATTAATCCTTAATGCTAAACACATTGAAAGGCGAAACACTACT 279
QY 52 ThrAlaIleGluAlaAlaSerSerAsp-----Ala 61
Db 280 GAAGCTGTGTGCTGCTACTGACAGAAAAGTCTTCAACAATACGCTAACACCAACCGT 339
QY 62 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGln-----SerGluGluAla 77
Db 340 GTTGACGGTGTAATGACTTACGACGATGCGACTAAGACCTTTTACAGTTACTGAAAACCA 399
QY 78 AlaValValLysAlaAsp-----Asn 84
Db 400 GAAGTGATCGATGCTCTGTAATTAACACCGCGTGACAACTTACAAAATCTGTTTATTAAT 459
QY 85 AlaAlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGlu 104
Db 460 GCTAAACATTTGAAGGCGAAACAACTACTAAAGCAGTAGACGCGAGAAACTCCAGAAAA 519
QY 105 Ala-----GluValValGlnSerAspAlaAlaAlaSerAspAlaTrp----- 118
Db 520 GCCTTCAAAACATAGCTTAACGACAAACGGTGTGTGTTGGTGTGTTGACTTATGATGCG 579
QY 118 ----- 118
Db 580 ACTAAGACTTTACGGTAACTGAAATGGTTACAGAGTTCTTCGAGGTGATCACAACACT 639
QY 119 -----GluLys-----AlaAlaThrProIle 125
Db 640 GAACAGAAAAAACCAAGCAAGTATCCCTCTTGTTCGTTAACTCCTGCAACTCAATT 699
QY 126 AlaLeuAspValLysLysThrLysAspThr-----LysProValVal 139
Db 700 GCTAAAGATGACGCTAAGAAAGACGATACTAAGAAAGAGAGATGCTAAAAAACCCAGAGCT 759
QY 140 LysLysGluGluArgGlnAsnValAsnThr 149
Db 760 AAGAAAGATGACGCTAAGAAAGCTGAAACT 789
RESULT 16
STRABP STRABP 969 bp DNA linear BCT 26-APR-1993
LOCUS Streptococcus canis (group G) albumin-binding protein gene, partial cds.
ACCESSION M95520
VERSION M95520.1 GI:153554
KEYWORDS albumin-binding protein.
SOURCE Streptococcus canis (group G) (individual isolate DG12) DNA.
ORGANISM Streptococcus canis
Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae; Streptococcus.
REFERENCE 1 (bases 1 to 969)
AUTHORS Sjobring,U.
TITLE Isolation and molecular characterization of a novel albumin-binding protein from group G streptococci
```

Pred. No.:	0.00155	Length:	1576
Score:	154.50	Matches:	80
Percent Similarity:	26.45%	Conservative:	16
Best Local Similarity:	22.04%	Mismatches:	62
Query Match:	19.88%	Indels:	206
DB:	1	Gaps:	7
US-09-847-539a-6 (1-159) x SG148PG (1-1576)			
QY	1	ValAspSerProIleGluGlnProAArgIleIleProAsnGlyThrLeuThrAsnLeu	20
Db	391	GTGACTCAACCAATCGAAGATACCCCAATTATTGCTAATGGTGAATTAACCTAATCTT	450
QY	21	LeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaIleAspGluLeu	40
Db	451	CTGGGAATTCAGACACAACTGGCTTTCGCTAATGAAGAGAGTGTACAGCTGATTG	510
QY	41	LysLysGlnAlaIleGluAspLysGluAlaThrAlaIleGluAlaSerSerAsp	60
Db	511	ACAGCAGCAGCGGTAGCCGATCTGTGCAGCAGCGAGCTGAAATGCTGGCGACGA	569
QY	61	AlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaIleVal	80
Db	570	GCTTGGGAAGCAGCGGCA---GCAGCAGATGCTCTAGCAAAAGCCAAAGCAGATGCCCTT	626
QY	81	LysAlaAspAsn-----	84
Db	627	AAAGATTCAACATAGATGAATTTTACGTGCATTACCTAAGACTGACACTTACAAATTA	686
QY	85	-----AlaAlaSerAspAlaLeuGluAlaLeuAla	94
Db	687	ATCCTTAATGGTAAACATTGAAAGCGAAACAACACTACTCAAGCTGTGTATGCTGCTACT	746
QY	95	AspGln-----	96
Db	747	GCAGAAAAAGCTTCAAAACAATACGCTACGACCAACGGTGTGACGGTGAATGGACTTAC	806
QY	96	-----	96
Db	807	GACGATCGGACTAAGACCTTTACAGTTACTTACGATGCGATGCGTCTGAA	866
QY	96	-----	96
Db	867	TTAACACCGCGGTGACAACTTACAAACTTGTATTAAATGGTAAACACTTGAAGCGGAA	926
QY	97	-----ThrAspAlaLeuGlnSerGluGluAlaGluVal-----GlnSer	110
Db	927	CAACTACTGAAGCTGTTGATGCTGCTACTGCAGAAAAAGTCTTCAAAACAATACGCTAAC	986
QY	111	AspAsnAlaAlaSerAspAlaTIp-----	118
Db	987	GACACGGTGTTCACGGTGAATGGACTTACGACGATGCGACTAAGACCTTTACAGTTACT	1046
QY	118	-----	118
Db	1047	GA AAAACAGAGTATGATGATGCGTCTGAATTAACACACCGCGTACAACTTACAACTT	1106
QY	118	-----	118
Db	1107	GTTATTAAATGGTAAACATTGAAAGCGGAAACAACACTACTAAGCAGTAGACGCGAGAACT	1166
QY	119	---GluLysAla-----	121
Db	1167	GCAGAAAAAGCCCTTCAAAACAATACGCTAACGACAACGGTGTGATGGTGTTCGACTTAT	1226
QY	121	-----	121
Db	1227	GATGATCGGACTAAGACCTTTACGGTAACTGAATAAGTTTACAGAGTTCTCTGCTGATGC	1286
QY	122	-----AlaThr	123
Db	1287	ACCAACTGAACCAAGAAAAACCAAGCAAGTATCCCTCTTTGTTCCGTTAACTCCTCGCAAC	1346

QY	123	rProIleAlaLeuAspValLysLysThrLysAspThr-----LysPr	137
Db	1347	TCCAAATGCTAAAGATCACGCTAAGAAAGACGATACTAAGAAAGACATGCTAAATAACC	1406
QY	137	oValValLysLysGluGluArgGlnAsnValAsnThrLeuProThrThrGlyGluGluSe	157
Db	1407	AGAAGCTAAGAAAGACGCTAAGAAAGCTGAAACTCTTCTACAACTGCTGAAGGANG	1466
QY	157	rAsnPro	159
Db	1467	CAACCCA	1473
RESULT 14			
SSCMLB1			
LOCUS		1380 bp	DNA linear BCT 12-JUN-1996
DEFINITION		Streptococcus sp.serotype C mbl1 gene for multiple ligand-binding protein 1.	
ACCESSION		X84989	
VERSION		GI:1403574	
KEYWORDS		mb1 gene; multiple ligand-binding protein.	
SOURCE		Streptococcus sp.	
ORGANISM		Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae; Streptococcus.	
REFERENCE		1 (bases 1 to 1380)	
AUTHORS		Talay, S.R.	
TITLE		Direct Submission	
JOURNAL		Submitted (23-FEB-1995) S.R. Talay, GBF Gesellschaft f. Biotechn. Forschung mbH, National Research Centre for Biotechn., Mascheroder Weg 1, D-38124 Braunschweig, FRG	
REFERENCE		2 (bases 1 to 1380)	
AUTHORS		Talay, S.R., Grammel, M.P. and Chhatwal, G.S.	
TITLE		Structure of a group C streptococcal protein that binds to fibrinogen, albumin and immunoglobulin G via overlapping modules	
JOURNAL		Biochem. J. 315 (Pt 2), 577-582 (1996)	
MEDLINE		96202013	
FEATURES		Location/Qualifiers	
source		1..1380	
		/organism="Streptococcus sp."	
		/strain="serotype C"	
		/isolate="C20"	
		/db_xref="taxon:1306"	
sig_peptide		61..168	
		/gene="mb1"	
gene		61..1380	
		/gene="mb1"	
CDS		61..1380	
		/gene="mb1"	
		/codon_start=1	
		/transl_table=11	
		/evidence=experimental	
		/product="multiple ligand-binding protein 1"	
		/protein_id="CAA59349.1"	
		/db_xref="GI:1403575"	
		/db_xref="SPTREMBL:Q55105"	
		/translation="MARKKTKQYSIRKLSVGAASVLVATGLIGGATVSAASEYSDRIT TGTSNGHITSSGIGSTIVGGLAKPASNIIGTGTAGTIGSGTGGTIVSGSGN TASGTVAGAGTTHDGSNTIASGVEEKDLKVEAELEANEALSTILDKTIEEKI SEYKKEIGIKELDHINNLGNKADQDLKITEKESLSDKDEALDERNKVLEASR TRNRDLEAAROKATEAELEAKVDKLEEKOVLEASRRKTRNRDLEAARDAKA TEAELEAKNELNQILEASRTRNRDLEAARDAKAVDAELAKLAKAEALKEOLAKQA QELEKLESKEKAPETQTPKEPEVPGKSPMPWTLTPATPTAKDKKTDVPRPAKAN MPTDVKKDEKLLPSTGETVNFPTAAGMAGMATAGVAVAGRRKEN"	
BASE COUNT	530 a	244 c	314 g 292 t
ORIGIN			
Alignment Scores:			
Pred. No.:	0.0023	Length:	1380
Score:	151.50	Matches:	65
Percent Similarity:	37.66%	Conservative:	25
Best Local Similarity:	27.20%	Mismatches:	46
Query Match:	19.50%	Indels:	103
DB:	1	Gaps:	10



```
Qy 21 LeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaAlaLeuAspGluLeu 40
|||||
Db 505 CTGGGAATTCAGACACACACTGCTTGGCTAATGAAGAGAGTGCTACAGCTGATTG 564
|||||
Qy 41 LysLysGlnAlaIleGluAspLysGluAlaThrThrAlaIleGluAlaAlaSerSerAsp 60
|||||
Db 565 ACAGCAGCAGCGGTACCGCATACTGTGCAGCAGCGCGCAGCTGAATGCTGGGGCAGCA 624
|||||
Qy 61 AlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaValVal 80
|||||
Db 625 GCTTGGGAAGCAGCGGCA---GCAGCAGATGCTCTAGCAAAAGCCAAAGCAGATGCCCTT 681
|||||
Qy 81 LysAlaAspAsn-----AlaAlaSerAspAlaLeuGluAlaLeu-----AlaAsp 95
|||||
Db 682 AAAGAAATTCACAATATGGAGTACGTACTATTACAGAATCTAATCAACAATGCCAAA 741
|||||
Qy 96 GlnThrAspAlaLeuGlnSerGluGluAlaGluValValGlnSerAspAsnAlaAla--- 114
|||||
Db 742 ACTGTTGAAGCGGTAAAGACCTTCAAGCACAAGTTGTTGAATCAGCGCAAGAAAGCGGT 801
|||||
Qy 115 ---SerAspAlaTrpGlu-----LysAlaAlaThrProIleAlaLeu 127
|||||
Db 802 ATTCAGAGCAACAGATGGCTTATCTGATTTCTTGAATCACAACACCTCTGCTGAAGAT 861
|||||
Qy 128 AspValLysLysThrLys-----AspThrLysProValValLysLysGlu 142
|||||
Db 862 ACTGTTAAATCAATTAAGTATTAGCTCAAGCTAAAGCTTCTAGCTAACAGAGAA 912
|||||
RESULT 12
LOCUS I08536 1950 bp linear PAT 02-DEC-1994
DEFINITION Sequence 2 from Patent WO 8705025.
ACCESSION I08536
VERSION I08536.1 GI:588754
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
BASE COUNT 706 a 324 c 397 g 523 t
ORIGIN
Alignment Scores:
Pred. No.: 7.29e-06 Length: 1950
Score: 186.00 Matches: 61
Percent Similarity: 51.59% Conservative: 20
Best Local Similarity: 38.85% Mismatches: 60
Query Match: 23.94% Indels: 16
DB: 6 Gaps: 6
US-09-847-539A-6 (1-159) x I08536 (1-1950)
Qy 1 ValAspSerProIleGluGlnProArgIlelleProAsnGlyGlyThrLeuThrAsnLeu 20
|||||
Db 677 GTTGATTCACCAATCGAAGATACCCCAATTAATTCGTAATGGTGTGTAATTAATCTT 736
|||||
Qy 21 LeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaAlaLeuAspGluLeu 40
|||||
Db 737 CTGGGAATTCAGACACACACTGCTTGGCTAATGAAGAGAGTGCTACAGCTGATTG 796
|||||
Qy 41 LysLysGlnAlaIleGluAspLysGluAlaThrThrAlaIleGluAlaAlaSerSerAsp 60
|||||
Db 797 ACAGCAGCAGCGGTACCGCATACTGTGCAGCAGCGCGCAGCTGAATGCTGGGGCAGCA 856
|||||
Qy 61 AlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaValVal 80
```

```
Db 857 GCTTGGGAAGCAGCGGCA---GCAGCAGATGCTCTAGCAAAAGCCAAAGCAGATGCCCTT 913
|||||
Qy 81 LysAlaAspAsn-----AlaAlaSerAspAlaLeuGluAlaLeu-----AlaAsp 95
|||||
Db 914 AAAGAAATTCACAATATGGAGTACGTACTATTACAGAATCTAATCAACAATGCCAAA 973
|||||
Qy 96 GlnThrAspAlaLeuGlnSerGluGluAlaGluValValGlnSerAspAsnAlaAla--- 114
|||||
Db 974 ACTGTTGAAGCGCATAAAGACCTTCAAGCACAAGTTGTTGAATCAGCGCAAGAAAGCGGT 1033
|||||
Qy 115 ---SerAspAlaTrpGlu-----LysAlaAlaThrProIleAlaLeu 127
|||||
Db 1034 ATTCAGAGCAACAGATGGCTTATCTGATTTCTTGAATCGCAACACCTCTGAAGAT 1093
|||||
Qy 128 AspValLysLysThrLys-----AspThrLysProValValLysLysGlu 142
|||||
Db 1094 ACTGTTAAATCAATTAAGTATTAGCTCAAGCTAAAGCTTCTAGCTAACAGAGAA 1144
|||||
RESULT 13
LOCUS SGI48PG 1576 bp DNA linear BCT 23-MAR-1993
DEFINITION Streptococcus G148 protein G' structural gene.
ACCESSION X53324
VERSION X53324.1 GI:288358
KEYWORDS Streptococcal Protein G.
SOURCE Streptococcus sp., 'group G'.
ORGANISM Streptococcus sp., 'group G'.
REFERENCE
AUTHORS Goward,C.R., Murphy,J.P., Atkinson,T. and Barstow,D.A.
TITLE Expression and purification of a truncated recombinant streptococcal protein G
JOURNAL Biochem. J. 267 (1), 171-177 (1990)
MEDLINE 90226312
FEATURES
source
1..1576
/organism="Streptococcus sp. 'group G'"
/strain="G148"
/db_xref="taxon:1320"
/clone="pspg3"
292..1262
/gene="Protein G'gene"
292..603
/gene="Protein G'gene"
note="Truncated gene"
/codon_start=1
/transl_table=11
/product="Protein G'"
/protein_id="CAA37409.1"
/db_xref="GI:288359"
/db_xref="SPTREMBL:054180"
/translation="MERKKVKYELRKSAGFLGSVAFLVGVSTVFAVDSIEDTPII
RNGELTNLLGNSETTLALRNEESATADLTAAAVADTVAAAAAENAGQGLGKQKQOQ
L"
705..1262
/gene="Protein G'gene"
note="Truncated gene; Start codon TTG not ATG"
/codon_start=1
/transl_table=11
/product="Protein G'"
/protein_id="CAA37410.1"
/db_xref="GI:581662"
/db_xref="SPTREMBL:054181"
/translation="MKGETTTEAVDAATAEKVFKQYANDNCVDGENTYDDATKTKFTVT
EKPEVIDASELTPAVTYYKLINGTKLGETTTEAVDAATAEKVFKQYANDNGVDGEW
TYDATKFTVTEKPEVIDASELTPAVTYYKLINGTKLGETTTRKAVDAETAEKAFK
QYANDNGVDGWTYDDATKTKFTVE"
BASE COUNT 572 a 273 c 327 g 404 t
ORIGIN
Alignment Scores:
```

ORGANISM	Unknown.		Location/Qualifiers	
	Unclassified.		1. .2456	
	REFERENCE 1 (bases 1 to 2384)		/organism="Streptococcus sp. G148"	
	AUTHORS Fahnstocck,S.R.		/strain="G148"	
	TITLE CLONED STREPTOCOCCAL GENES ENCODING PROTEIN G AND THEIR USE TO CONSTRUCT RECOMBINANT MICROORGANISMS TO PRODUCE PROTEIN G		/db_xref="taxon:1324"	
JOURNAL	Patent: WO 8705025-A 4 27-AUG-1987;		/clone="pspg1 and pspg8"	
	FEATURES		335. .340	
	Location/Qualifiers		/note="pot. ribosome binding site"	
	source		346. .444	
	1. .2384		346. .2127	
BASE COUNT	868 a 410 c 480 g 626 t		/note="G preprotein (AA -33 to 560)"	
	ORIGIN		/codon_start=1	
	Alignment Scores:		/transl_table=11	
	Pred. No.:		/protein_id="CAA29540.1"	
	Score:		/db_xref="GI:47085"	
Percent Similarity:	187.00		/db_xref="SWISS-PROT:P19909"	
	51.59%		/translation="MEKEKKVKYFLRKSAFGLASVSAFLVGVSTVFAVDSPIEDTPII	
	38.85%		RNGELTNLNGSETTLALRNEESATADLTAAAVADTVAAAAAENAGAAWAAAAAD	
	24.16		ALAKAKADALKKEFNKGVSDYYKLNINNAKTVEGVDLQAOVVESAKKARISEATDGL	
	6		SDPLKSOPTAEDTKVKSIELAEAKVLANRELDKYGVSVDYHKNLINNAKTVEGVKDLQAO	
DB:	2384		VYESAKKARISEATDGLSDFLSQTPAEDTVKYSIELAEAKVLANRELDKYGVSDDYKN	
	108537 (1-2384)		LINNAKTVEGVKALIDEILUALPKTDYTKLIINGKTLKGTTTEAVDAATAEKVPKQY	
	QY 1 ValAspSerProfileGluInProArgIleIleProAsnGlyGlyThrLeuThrAsnLeu 20		ANDNGVGEWYDDATKTFTVTEKPEVIDASELTPAVTYTKLVINGTKTLKGTTTEAV	
	Db 676 GTTGACTCACCATCGAAGATACCCCAATATTTCGTAATGGTGGTCAATTAACATACTT 735		DAATAEKVKQYANDNGVGEWYDDATKTFTVTEKPEVIDASELTPAVTYTKLVINGTKTLKGTTTEAV	
	QY 21 LeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaIleAspGluLeu 40		EPEKPEASIPLVPLTPATPIAKDDAKKDDTKEDAKKPEAKKEDAKKAETLPTTGSGS	
DB:	736 CTGGGAATTCAGAGACAACACTGGCTTTGCGTAATGAAGAGAGTGTACAGCTGATTG 795		mat_peptide	
	QY 41 LysLysGlnAlaIleGluAspLysGluAlaThrThrAlaIleGluAlaAlaSerSerAsp 60		repeat_region	
	Db 796 ACAGCAGCAGCGGTACCGATCTGTGGCAGCAGCGCGCTGAATGCTGGGCGACGA 855		repeat_region	
	QY 61 AlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVal 80		repeat_region	
	Db 856 GCTTGGGAAGCAGCGCGCA---GCAGCAGATGCTCTAGCAAAAGCCAAAGCAGATGCCCTT 912		repeat_region	
QY	81 LysAlaAspAsn-----AlaAlaSerAspAlaLeuGluAlaLeu-----AlaAsp 95		repeat_region	
	Db 913 AAGAATTCAACAATAATATGGAGTAAGTACTATTACAAGAAATCAACAATGCCAA 972		repeat_region	
	QY 96 GlnThrAspAlaLeuGlnSerGluGluAlaGluValValGlnSerAspAsnAlaAla--- 114		repeat_region	
	Db 973 ACTGTTGAAGCGGTAAGACCTTCAAGCACAAGTGTGTAATCAGCGAAGAAAGCGCGT 1032		repeat_region	
	QY 115 ---SerAspAlaTrpGlu-----LysAlaAlaThrProIleAlaLeu 127		repeat_region	
Db	1033 ATTTCAAGAACACACATGCTTATCTGATTTCTTGAATCACAACACCTGCTGAAGAT 1092		repeat_region	
	QY 128 AspValLysLysThrLys-----AspThrLysProValValLysLysGlu 142		repeat_region	
	Db 1093 ACTGTTAAATCAATTGAATTAGCTGAAGCTGAAGTCTTAGCTAACAGAGAA 1143		repeat_region	
	RESULT 11		repeat_unit	
SGPROTG	SGPROTG		repeat_unit	
	LOCUS		BASE COUNT	
	DEFINITION		ORIGIN	
	ACCESSION		907 a 412 c 491 g 646 t	
	VERSION		X06173.1 M28331	
KEYWORDS	G protein; IgG receptor.		Alignment Scores:	
	Streptococcus sp. G148.		Pred. No.:	
	Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;		Score:	
	Streptococcus.		Percent Similarity:	
	1 (bases 1 to 2456)		Best Local Similarity:	
AUTHORS	Olsson,A., Eliasson,M., Guss,B., Nilsson,B., Hellman,U.,		Query Match:	
	Lindberg,M. and Uhlen,M.		DB:	
	Structure and evolution of the repetitive gene encoding		US-09-847-539A-6 (1-159) x SGPROTG (1-2456)	
	streptococcal protein G		QY 1 ValAspSerProfileGluGlnProArgIleIleProAsnGlyGlyThrLeuThrAsnLeu 20	
	Eur. J. Biochem. 168 (2), 319-324 (1987)		Db 445 GTTGACTCACCATCGAAGATACCCCAATATTTCGTAATGGTGGTCAATTAACATACTT 504	
JOURNAL	MEDLINE		Alignment Scores:	
	88029445		Pred. No.:	
			Score:	
			Percent Similarity:	
			Best Local Similarity:	
MEDLINE			Query Match:	
			DB:	
			US-09-847-539A-6 (1-159) x SGPROTG (1-2456)	
			QY 1 ValAspSerProfileGluGlnProArgIleIleProAsnGlyGlyThrLeuThrAsnLeu 20	
			Db 445 GTTGACTCACCATCGAAGATACCCCAATATTTCGTAATGGTGGTCAATTAACATACTT 504	

```

Db 797 ACAGCAGCAGCGGTAGCCGACTACTGTGGCAGCAGCGGCGAGCTGAATAATCTGCTGGGCAGCA 856
Qy 61 AlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVal 80
   ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||||
Db 857 GCTTGGGAAGCAGCGGCA---GCAGCAGATGCTCTAGCAAAAGCCAAACGACGATGCCCTT 913
Qy 81 LysAlaAspAsn-----AlaAlaSerAspAlaLeuGluAlaLeu-----AlaAsp 95
   ||| ||| ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 914 AAAGAATTCAACAATAATATGAGTAAGTACTATTACAGAATCTTAATCAACAATGCCAAA 973
Qy 96 GlnThrAspAlaLeuGlnSerGluGluAlaGluValValGlnSerAspAsnAlaAla--- 114
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 974 ACTGTTGAAGGATAAAGACCTTCAAGCACACAGTGTGTAATCAGCGAAGAAAGCGCT 1033
Qy 115 ---SerAspAlaTrpGlu-----LysAlaAlaThrProIleAlaLeu 127
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
Db 1034 ATTTCAAGAAGCACAGCTATCTGATTTCTTGAATCGCAACACCTGCTGAAGAT 1093
Qy 128 AspValLysLysThrLys-----AspThrLysProValValLysLysGlu 142
   ||||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
Db 1094 ACTGTTAAATCAATTGAATAGCTGAAGCTAAAGTCTTAGCTAACACAGAA 1144

RESULT 9
SGSPG
LOCUS
DEFINITION Streptococcus sp. spg gene for protein G. linear BCT 09-OCT-1997
ACCESSION Y00428
VERSION Y00428.1 GI:47100
KEYWORDS G protein; IgG binding protein; spg gene.
SOURCE Streptococcus sp. GX7805.
ORGANISM Streptococcus.
REFERENCE 1 (bases 1 to 2384)
AUTHORS Fahnestock, S.R.
TITLE Direct Submission
JOURNAL Submitted (12-OCT-1987) Fahnestock S.R., Genex Corp., 16020
Industrial Dr., Gaithersburg, MD 20877, USA
REFERENCE 2 (bases 1 to 2384)
AUTHORS Filpula, D., Alexander, P. and Fahnestock, S.R.
TITLE Nucleotide sequence of the protein G gene from Streptococcus
GX7805, and comparison to previously reported sequences
JOURNAL Nucleic Acids Res. 15 (17), 7210 (1987)
MEDLINE 88015586
FEATURES
    source
        1..2384
        /organism="Streptococcus sp. GX7805"
        /strain="GX7805"
        /db_xref="taxon:1325"
        /db_xref="taxon:1325"
        577..2358
        /gene="spg"
        /gene="spg"
        /gene="spg"
        /gene="spg"
        /codon_start=1
        /transl_table=11
        /product="protein G precursor"
        /protein_id="CAA68489.1"
        /db_xref="GI:47101"
        /db_xref="SWISS-PROT:P19909"
        /translation="MEKEKKVKYFLRKSAFLGSLVSAFLVGSTVFVAVDSPIEDTPII
        RRGELTNLGNSETTLALRNEESATADLTAAAVADTVAAAAENAGAAWEAAAAAD
        ALAKADALKEFNKYGVSDYYKNLINNAKTVEGVKDLQAOVSEAKARISEATDGL
        SDFLSQTPAEDTVKSIELAEKVLNRELDKYGVSDYHKNLINNAKTVEGVKDLQAO
        VSEAKARISEATDGLSDFLSQTPAEDTVKSIELAEKVLNRELDKYGVSDYHKN
        LINNAKTVEGVKDLSEILALPQTDPTKYLINGKTLLGETTTEAVDAATAEKVEKY
        ANDNGVDEMTYDDATKFTVTEKPEVIDASELTPAVTYKLVLINGKTLKGETTEAV
        DAATAEKVEKYANDNGVDEMTYDDATKFTVTEKPEVIDASELTPAVTYKLVLING
        KTLKGETTTEAVDAATAEKVEKYANDNGVDEMTYDDATKFTVTEKPEVIDASELTP
        AVTYKLVLINGKTLKGETTTEAVDAATAEKVEKYANDNGVDEMTYDDATKFTVTEK
        PEVIDASELTPAVTYKLVLINGKTLKGETTTEAVDAATAEKVEKYANDNGVDEMTY
        DDATKFTVTEKPEVIDASELTPAVTYKLVLINGKTLKGETTTEAVDAATAEKVEKY
        NPFFTAAALAVMAGALAVASKRED"
        676..2355
    mat_peptide

```

```

/gene="spg"
/product="protein G"
886..996
/Note="A1 repeated sequence"
986..1435
/Note="A2 repeated sequence"
1111..1221
/Note="A3 repeated sequence"
1354..1446
/Note="B1 repeated sequence"
1483..1647
/Note="B2 repeated sequence"
1499..1918
/Note="B3 repeated sequence"
1693..1857
/Note="B4 repeated sequence"
1903..2067
/Note="B5 repeated sequence"
BASE COUNT 868 a 408 c 482 g 626 t
ORIGIN
Alignment Scores:
Pred. No.: 7,51e-06 Length: 2384
Score: 187.00 Matches: 61
Percent Similarity: 51.59% Conservative: 20
Best Local Similarity: 38.85% Mismatches: 60
Query Match: 24.07% Indels: 16
DB: 1 Gaps: 6
US-09-847-539a-6 (1-159) x SGSPG (1-2384)
Qy 1 ValAspSerProIleGluGlnProArgIleIleProAsnGlyGlyThrLeuThrAsnLeu 20
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 676 GTTGACTCAACCAATCGAAGTACCCCAATATTTCGTAATGGTGGTAATACTAATCTT 735
Qy 21 LeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaAlaLeuGluLeu 40
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 736 CTGGGAATTCAGAGACAACACTGGCTTTCGTAATGAAGAGAGTCTACAGCTGATTG 795
Qy 41 LysLysGluAlaAlaGluAspLysGluAlaThrAlaAlaGluAlaAlaSerSerAsp 60
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 796 ACAGCAGCAGCGGTAGCCGACTACTGTGGCAGCAGCGGCGAGCTGAATAATCTGCGGCGCA 855
Qy 61 AlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVal 80
   ||| ||||| ||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||||
Db 856 GCTTGGGAAGCAGCGGCA---GCAGCAGATGCTCTAGCAAAAGCCAAACGACGATGCCCTT 912
Qy 81 LysAlaAspAsn-----AlaAlaSerAspAlaLeuGluAlaLeu-----AlaAsp 95
   ||| ||| ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 913 AAAGAATTCAACAATAATATGAGTAAGTACTATTACAGAATCTTAATCAACAATGCCAAA 972
Qy 96 GlnThrAspAlaLeuGlnSerGluGluAlaGluValValGlnSerAspAsnAlaAla--- 114
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 973 ACTGTTGAAGGCGTAAAGACCTTCAAGCACACAGTGTGTAATCAGCGAAGAAAGCGCT 1032
Qy 115 ---SerAspAlaTrpGlu-----LysAlaAlaThrProIleAlaLeu 127
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
Db 1033 ATTTCAAGAAGCACAGTATCTGATTTCTTGAATCGCAACACCTGCTGAAGAT 1092
Qy 128 AspValLysLysThrLys-----AspThrLysProValValLysLysGlu 142
   ||||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
Db 1093 ACTGTTAAATCAATTGAATAGCTGAAGCTAAAGTCTTAGCTAACACAGAA 1143
RESULT 10
LOCUS
DEFINITION Sequence 4 from Patent WO 8705025. linear PAT 02-DEC-1994
ACCESSION I08537
VERSION I08537.1 GI:588755
KEYWORDS Unknown.

```

```

|||||
Db 481 GCTCGGCTGTAAAGCGGATAACGCTGCTAGTGACACTTTAGAACGATTTGGCGGATCAA 540
|||||
Qy 97 ThrAspAlaLeuGlnSerGluGluAlaGluValValGlnSerAspAsnAlaAlaSerAsp 116
|||||
Db 541 ACAGACGCTTTACAAATCAGAGAAGAGCTGAAGTAGTTCAATCAGATAACGCTGCTAGTGAC 600
|||||
Qy 117 AlaTrpGluLysAlaAlaThrProIleAlaLeuAspValLysLysThrLysAspThrLys 136
|||||
Db 601 GCCTGGGGAAGAGCAGCACTCCAAATCGCTTTAGATGTTAAGAAAACCTAAAGATACAAAA 660
|||||
Qy 137 ProValValLysLysGluGluAlaGlnAsnValAsnThrLeuProThrGlyGlu 155
|||||
Db 661 CCTGTACTTAAAGAAAGAAAGACAAACGTTAATACCTTCCTACAACTGGTGAA 717
|||||
RESULT 7
STRSPGIGP 1950 bp DNA linear BCT 23-DEC-1993
LOCUS Streptococcus sp. (Lancefield group G) spg gene encoding an
DEFINITION Immunoglobulin G binding protein.
ACCESSION M13825.1 GI:153822
VERSION constant region; Immunoglobulin binding protein.
KEYWORDS Streptococcus sp. (Lancefield group G; strain GX7809; DNA, clone
SOURCE MGX4547.
ORGANISM Streptococcus sp.
Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
Streptococcus.
REFERENCE 1 (bases 1 to 1950)
AUTHORS Fahnestock,S.R., Alexander,P., Nagle,J. and Filpula,D.
TITLE Gene for an immunoglobulin-binding protein from a group G
streptococcus
JOURNAL J. Bacteriol. 167, 870-880 (1986)
MEDLINE 86304178
COMMENT Draft entry and computer-readable sequence of [1] kindly provided
by S.R.Fahnestock, 07-NOV-1986. The -35 and -10 regions are
located at positions 465-470 and 487-492 respectively and an SD
sequence at 565-570.
FEATURES
source 1..1950
/organism="Streptococcus sp."
/db_xref="taxon:1306"
sig_peptide 578..676
/notes="IG binding protein signal peptide"
CDS 578..1924
/notes="IG binding protein"
/codon_start=1
/transl_table=11
/protein_id="AAA03664.1"
/db_xref="GI:153823"
/translation="MEKEKKVKYFLRKSAPGLASVSAFLVSGTVFAVDSPIEDTPII
RNGELTNLLGNSETFLALRNESATADLTAAVADTVAAAEENAGAAWEAAAAAD
ALAKADALKEPNKYGVSDYYKLNINNAKTVEGIDKQAVVSESARKARISATDGL
SDFKSTPAEDTVKYSIELAEAKYLANRELDKYGVSDYHKLNINNAKTVEGVKVELIDE
ILAALEPKTDTYKILNGKTLKGETTTTAVDAATAEKVKQYANDNGVDGEMTVYDDATK
TFVTVEKPEVIDASELTPAVTYKLVINGKTLKGETTTTAVDAETAERKAPKVANDNG
VDGWTYDDATKTFTVEMTVPGDAPTEPEKPEASIPVPLTPATPTAKDDAKDD
TKEDAKDPEAKDDAKKAETLPTTGEGSNPFTTAAALVAGAGALAVASKRRKD"
mat_peptide 677..1921
/product="IgG binding protein mature peptide"
BASE COUNT 706 a 323 c 398 g 523 t
ORIGIN 1 bp upstream of HindIII site.
Alignment Scores:
Pred. No.: 5.11e-06 Length: 1950
Score: 188.00 Matches: 61
Percent Similarity: 51.59% Conservative: 20
Best Local Similarity: 38.85% Mismatches: 60
Query Match: 24.20% Indels: 16
DB: 1 Gaps: 6
US-09-847-539A-6 (1-159) x STRSPGIGP (1-1950)

```

```

Qy 1 ValAspSerProIleGluGlnProArgIleProAsnGlyGlyThrLeuThrAsnLeu 20
|||||
Db 677 GTTGATTACCAATCGAAGATACCCCAATTATTCGTATGCTGCTGAATTAACCTAATCTT 736
|||||
Qy 21 LeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaIleAspGluLeu 40
|||||
Db 737 CTGGGAATTCAGAGACAACACTGGCTTGGCTAATGAAGAGAGTGCTACAGCTGATTG 796
|||||
Qy 41 LysLysGlnAlaIleGluAspLysGluAlaThrThrAlaIleGluAlaAlaSerSerAsp 60
|||||
Db 797 ACAGCAGCAGCGGTAGCCGACTACTGTGGCAGCAGCGGCTGAAATGCTGGGGCAGCA 856
|||||
Qy 61 AlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVal 80
|||||
Db 857 GCCTGGGAAGCAGCGGCA---GCAGCAGATGCTCTAGCAANAGCAAGCAGATGCCCTT 913
|||||
Qy 81 LysAlaAspAsn-----AlaAlaSerAspAlaLeuGluAlaLeu-----AlaAsp 95
|||||
Db 914 AAGCAATTCAACAATAATATGGAGTAAGTGACTATTACAAGAACTAATCAACAATGCCAA 973
|||||
Qy 96 GlnThrAspAlaLeuGlnSerGluGluAlaGluValValGlnSerAspAsnAlaAla--- 114
|||||
Db 974 ACTGTTGAAGGCATAAAAGACCTTTCAAGCACAAAGTTGTTGAATCAGCGAAGAAAGCCGT 1033
|||||
Qy 115 ---SerAspAlaTrpGlu-----LysAlaAlaThrProIleAlaLeu 127
|||||
Db 1034 ATTTCAAGAACGACAGATGGCTTATCTGATTTCTTGAATTCGCAACACCTGCTGAAGAT 1093
|||||
Qy 128 AspValLysLysThrLys-----AspThrLysProValValLysLysGlu 142
|||||
Db 1094 ACTGTTAAATCAATTGAATTAGCTGAAGCTAAGCTTAGCTAACACAGAA 1144
|||||
RESULT 8
LOCUS I09115
DEFINITION Sequence 23 from Patent WO 8810306.
ACCESSION I09115
VERSION I09115.1 GI:598187
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1950)
AUTHORS Fahnestock,S.R.
TITLE CLONED PROTEIN G VARIANT GENES AND THE PROTEIN G VARIANTS EXPRESSED
JOURNAL Patent: WO 8810306-A 23 DEC-1988;
FEATURES
source 1..1950
/organism="unknown"
BASE COUNT 706 a 323 c 398 g 523 t
ORIGIN
Alignment Scores:
Pred. No.: 5.11e-06 Length: 1950
Score: 188.00 Matches: 61
Percent Similarity: 51.59% Conservative: 20
Best Local Similarity: 38.85% Mismatches: 60
Query Match: 24.20% Indels: 16
DB: 1 Gaps: 6
US-09-847-539A-6 (1-159) x I09115 (1-1950)

```

```

Qy 1 ValAspSerProIleGluGlnProArgIleProAsnGlyGlyThrLeuThrAsnLeu 20
|||||
Db 677 GTTGATTACCAATCGAAGATACCCCAATTATTCGTATGCTGCTGAATTAACCTAATCTT 736
|||||
Qy 21 LeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaIleAspGluLeu 40
|||||
Db 737 CTGGGAATTCAGAGACAACACTGGCTTGGCTAATGAAGAGAGTGCTACAGCTGATTG 796
|||||
Qy 41 LysLysGlnAlaIleGluAspLysGluAlaThrThrAlaIleGluAlaAlaSerSerAsp 60
|||||

```



AUTHORS Rasmussen,M., Muller,H.P. and Bjorck,L.  
TITLE Protein GRAB of streptococcus pyogenes regulates proteolysis at the bacterial surface by binding alpha2-macroglobulin  
J. Biol. Chem. 274 (22), 15336-15344 (1999)  
JOURNAL 99269061  
MEDLINE 10336419  
PUBMED  
REFERENCE 2 (bases 1 to 469)  
AUTHORS Rasmussen,M., Muller,H.P. and Bjorck,L.  
TITLE Direct Submission  
JOURNAL Submitted (28-JAN-1999) Cell and Molecular Biology, Molecular Pathogenesis, Soelvegatan 39, Lund 221 00, Sweden  
FEATURES  
source  
1. .469  
/organism="Streptococcus pyogenes"  
/strain="AP1"  
/db\_xref="taxon:1314"  
1. .>469  
/product="GRAB"  
1. .174  
/note="Region: alpha2-macroglobulin-binding A domain"  
<1. .>469  
/note="grab"  
/codon\_start=1  
/transl\_table=11  
/product="GRAB precursor"  
/protein\_id="AAD26339.1"  
/db\_xref="GI:4589081"  
/translation="VDSPIEQPRIIPNGGTLNLGNAPKIALRNEERAIDELKKQA  
IEDKEATTAEASDADALADQDLOSEAAVVKADNAASDALEALADQDLOQ  
EEAVVQSDNAASDAWEKAATPIALDVKKTKDKPVVKKERQNVNTLPTTGE"  
175. .342  
misc\_feature  
/note="Region: repeat motif"  
misc\_feature 343. .469  
/note="Region: cell-wall attachment"  
BASE COUNT 178 a 90 c 100 g 101 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 3.39e-50 Length: 469  
Score: 734.00 Matches: 155  
Percent Similarity: 99.36% Conservative: 0  
Best Local Similarity: 99.36% Mismatches: 1  
Query Match: 97.04% Indels: 0  
DB: 1 Gaps: 0  
US-09-847-539A-6 (1-159) x AF124400 (1-469)  
QY 1 ValAspSerProIleGluGlnProArgIleIleProAsnGlyGlyThrLeuThrAsnLeu 20  
|||||  
Db 1 GTCGACTCACCTATCGAACAGCGCTAGCAATATTCCAAATGCGCGAACCTTAATTAATCTT 60  
QY 21 LeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaIleAspGluLeu 40  
|||||  
Db 61 CTTGGCAATGCTCCAGAAACCTGGCATTACGTAATGAAGAAAGAGCCATTGATGAATTA 120  
QY 41 LysLysGlnAlaIleGluAspLysGluAlaThrAlaIleGluAlaAlaSerSerAsp 60  
|||||  
Db 121 AAAAACACAGCTATTGAGGATAGGAGCTTACCACAGCTATAGAAGCAGCAAGTTCAGAT 180  
QY 61 AlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVal 80  
|||||  
Db 181 GCCTTAGAGCATTAGCGGATCAACAGACGCTTTACAATCACAAGAGCTGCGGTGTT 240  
QY 81 LysAlaAspAsnAlaAlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeu 100  
Db 241 AAGCGCGATCAACGCTGCTAGTACGCTTAGAAGCATTGGCGGATCAACAGACGCTTTA 300  
QY 101 GlnSerGluGluAlaGluValValGlnSerAspAsnAlaAlaSerAspAlaTtpGluLys 120  
Db 301 CAATCAGAAGAGCTGAGTAGTTCATCATCAGATACGCTCTAGTCAGCGCTGGGAAAA 360  
QY 121 AlaAlaThrProIleAlaLeuAspValLysLysThrLysAspThrLysProValValLys 140  
|||||

Db 361 GCAGCAACTCCAATCGCTTTAGATGTTTAAAGAACTAAGATACAAACCTGTAGTTAAA 420  
QY 141 LysCluclArgGlnAsnValAsnThrLeuProThrThrGlyGluGlu 156  
|||||  
Db 421 AAAGAAGAAAGACAAACGTTAATACCTCTCTACAACTGGTGAAGAG 468  
RESULT 5  
AF124403 804 bp DNA linear BCT 14-AUG-2000  
LOCUS Streptococcus pyogenes strain AP49 GRAB precursor, gene, partial  
DEFINITION cds.  
ACCESSION AF124403  
VERSION AF124403.1 GI:4589086  
KEYWORDS Streptococcus pyogenes.  
SOURCE Streptococcus pyogenes  
ORGANISM Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
Streptococcus.  
REFERENCE 1 (bases 1 to 804)  
AUTHORS Rasmussen,M., Muller,H.P. and Bjorck,L.  
TITLE Protein GRAB of streptococcus pyogenes regulates proteolysis at the bacterial surface by binding alpha2-macroglobulin  
J. Biol. Chem. 274 (22), 15336-15344 (1999)  
JOURNAL 99269061  
MEDLINE 10336419  
PUBMED  
REFERENCE 2 (bases 1 to 804)  
AUTHORS Rasmussen,M., Muller,H.P. and Bjorck,L.  
TITLE Direct Submission  
JOURNAL Submitted (28-JAN-1999) Cell and Molecular Biology, Molecular Pathogenesis, Soelvegatan 39, Lund 221 00, Sweden  
FEATURES  
source  
1. .804  
/organism="Streptococcus pyogenes"  
/strain="AP49"  
/db\_xref="taxon:1314"  
1. .>804  
/product="GRAB"  
1. .174  
/note="Region: alpha2-macroglobulin-binding A domain"  
<1. .>804  
/note="grab"  
/codon\_start=1  
/transl\_table=11  
/product="GRAB precursor"  
/protein\_id="AAD26342.1"  
/db\_xref="GI:4589087"  
/translation="VDSPIEQPRIIPNGGTLNLGNAPKIALRNEERAIDELKKQA  
IEDKEATTAEASDADALADQDLOSEAAVVKADNAASDALEALADQDLOQ  
EEAVVQSDNAASDALEALADQDLOSEAAVVKADNAASDALEALADQDLOSEAE  
ASVVKADNAASDALEALADQDLOSEAAVVKADNAASDALEALADQDLOSEAE  
VVQSDNAASDAWEKAATPIALDVKKTKDKPVVKKERQNVNTLPTTGE"  
175. .678  
misc\_feature  
/note="Region: repeat motif"  
misc\_feature 679. .804  
/note="Region: cell-wall attachment"  
BASE COUNT 284 a 161 c 185 g 174 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 2.44e-45 Length: 804  
Score: 694.00 Matches: 156  
Percent Similarity: 58.21% Conservative: 0  
Best Local Similarity: 58.21% Mismatches: 0  
Query Match: 89.32% Indels: 112  
DB: 1 Gaps: 1  
US-09-847-539A-6 (1-159) x AF124403 (1-804)  
QY 1 ValAspSerProIleGluGlnProArgIleIleProAsnGlyGlyThrLeuThrAsnLeu 20  
|||||  
Db 1 GTTGACTCACCTATCGAACAGCGCTCGAATATTCCAAATGCGGAGACCTTAACCTAATCTT 60  
QY 21 LeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaIleAspGluLeu 40  
|||||

```
Pred. No.: 1.35e-50 Length: 10029
Score: 777.00 Matches: 159
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 1 Gaps: 0

US-09-847-539A-6 (1-159) x AF006573 (1-10029)

Qy 1 ValAspSerProIleGluGlnProArgIleLeuProAsnGlyGlyThrLeuThrAsnLeu 20
Db 6482 GTTGACTCACCTATCGACAGCGCTCGAATTATTCCTCAATGGCGAACCTTAACCTATCTT 6423

Qy 21 LeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluAlaIleAspGluLeu 40
Db 6422 CTTGGCAATGCTCCAGAAAACCTGGCATTAATGAAGAAGAGCCATTGATGAATTA 6363

Qy 41 LysLysGlnAlaIleGluAspLysGluAlaThrAlaIleGluAlaSerSerAsp 60
Db 6362 AAAAACAAGCTATTGAGGATAAAGAAGCTACGACAGCTATTAGAAGCAGCAAGTTTCAGAT 6303

Qy 61 AlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVal 80
Db 6302 GCCTTAGAGCATTAAGCGATCAACAGACGCTTACATCATCAGAAGAGCTCGCGTTGTT 6243

Qy 81 LysAlaAspAsnAlaAlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeu 100
Db 6242 AAAGCGGATAAGCTGCTAGTGACGCTTAGAAGCATTTGGCGGATCAACAGACGCTTTA 6183

Qy 101 GlnSerGluGluAlaGluAlaValGlnSerAspAsnAlaAlaSerAspAlaTrpGluLys 120
Db 6182 CATTGAGAAGAGCTGAAGTAGTCAATCAGATAACGCTGCTAGTGACGCTGGGAAAAA 6123

Qy 121 AlaAlaThrProIleAlaLeuAspValLysLysThrLysAspThrLysProValValLys 140
Db 6122 CGACGAACCTCAATCGCTTTAGATGTTAAGAAACCTAAGATACAAAACCTCTAGTTAA 6063

Qy 141 LysGluGluArgGlnAsnValAsnThrLeuProThrThrGlyGluGluSerAsnPro 159
Db 6062 AAAGAAGAAGAACAAACCTTAATACCTTCTACAACTGGTGAAGAGTCTTAACCCA 6006

RESULT 3
AF124401 AF124401 468 bp DNA linear BCT 14-AUG-2000
LOCUS Streptococcus pyogenes strain KTL3 GRAB precursor, gene, partial
DEFINITION cds.
ACCESSION AF124401
VERSION AF124400.1 GI:4589082
KEYWORDS Streptococcus pyogenes.
SOURCE Streptococcus pyogenes.
ORGANISM Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
REFERENCE 1 (bases 1 to 468)
AUTHORS Rasmussen,M., Muller,H.P. and Bjorck,L.
TITLE Protein GRAB of streptococcus pyogenes regulates proteolysis at the
JOURNAL bacterial surface by binding alpha2-macroglobulin
MEDLINE J. Biol. Chem. 274 (22), 15336-15344 (1999)
PUBMED 99269061
REFERENCE 2 (bases 1 to 468)
AUTHORS Rasmussen,M., Muller,H.P. and Bjorck,L.
TITLE Direct Submission
JOURNAL Submitted (28-JAN-1999) Cell and Molecular Biology, Molecular
Pathogenesis, Soelvestatan 39, Lund 221 00, Sweden
FEATURES
source Location/Qualifiers
1..468
/organism="Streptococcus pyogenes"
/strain="KTL3"
/db_xref="taxon:1314"
mat_peptide 1..>468
misc_feature 1..174

/region: alpha2-macroglobulin-binding A domain"
<1..>468
/feature="grab"
/codon_start=1
/transl_table=11
/product="GRAB precursor"
/protein_id="AAD26340.1"
/db_xref="gi:4589082"
/translation="VDSPIEQRIIPNGTTLNLGNAPKALRNERAIDEIKKQA
IEDKEATTAIEAASDADLEADOTDALOSEAAVVKADNAASDALALADQTDALOS
EEAEVQSDNAASDAWEKAATPIALDVKKTKDTKPVVKKKEERQNVNLTPTTGE"
175..342
/misc_feature /region: repeat motif"
343..468
/misc_feature /region: cell-wall attachment"
BASE COUNT 178 a 91 c 100 g 99 t
ORIGIN
Alignment Scores: 1.17e-50 Length: 468
Pred. No.: 760.00 Matches: 156
Score: 100.00% Conservative: 0
Percent Similarity: 100.00% Mismatches: 0
Best Local Similarity: 97.81% Indels: 0
Query Match: 1 Gaps: 0
DB:
US-09-847-539A-6 (1-159) x AF124401 (1-468)
Qy 1 ValAspSerProIleGluGlnProArgIleLeuProAsnGlyGlyThrLeuThrAsnLeu 20
Db 1 GTTGGACTCACCTATCGACAGCGCTCGAATTATTCCTCAATGGCGAACCTTAACCTATCTT 60
Qy 21 LeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluAlaIleAspGluLeu 40
Db 61 CTTGGCAATGCTCCAGAAAACCTGGCATTAATGAAGAAGAGCCATTGATGAATTA 120
Qy 41 LysLysGlnAlaIleGluAspLysGluAlaThrAlaIleGluAlaSerSerAsp 60
Db 121 AAAAACAAGCTATTGAGGATAAAGAAGCTACGACAGCTATTAGAAGCAGCAAGTTTCAGAT 180
Qy 61 AlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVal 80
Db 181 GCCTTAGAAGCATTAAGCGGATCAACAGACGCTTTACATCATCAGAAGAGCTCGGTTGTT 240
Qy 81 LysAlaAspAsnAlaAlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeu 100
Db 241 AAAGCGGATAACGCTGCTAGTGACGCTTAGAAGCATTTGGCGGATCAACAGACGCTTTA 300
Qy 101 GlnSerGluGluAlaGluAlaValGlnSerAspAsnAlaAlaSerAspAlaTrpGluLys 120
Db 301 CAATCAGAAGAGAGCTCAAGTAGTTCATCAATCAGATAACGCTGCTAGTGACGCTGGGAAAA 360
Qy 121 AlaAlaThrProIleAlaLeuAspValLysLysThrLysAspThrLysProValValLys 140
Db 361 CGAGCAACTCCCAATCGCTTAGATGTTAAGAAACCTAAGATACAAAACCTCTAGTTAAA 420
Qy 141 LysGluGluArgGlnAsnValAsnThrLeuProThrThrGlyGluGlu 156
Db 421 AAAGAAGAAGAACAAACCTTAATACCTTCTACAACTGGTGAAGAG 468

RESULT 4
AF124400 AF124400 469 bp DNA linear BCT 14-AUG-2000
LOCUS Streptococcus pyogenes strain AP1 GRAB precursor, gene, partial
DEFINITION cds.
ACCESSION AF124400
VERSION AF124400.1 GI:4589080
KEYWORDS Streptococcus pyogenes.
SOURCE Streptococcus pyogenes.
ORGANISM Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
REFERENCE 1 (bases 1 to 469)
```

```

CDS
complement(619..1911)
/gene="aroA"
/EC_number="2.5.1.19"
/notes="Best Blastp hit = sp|P43905|AROA_LACLA
3-PHOSPHOSHIMKIMATE 1-CARBOXYVINYLTRANSFERASE
(SYNTHASE) (EPSP) >gi|107524|pir||S52580
3-phosphoshikimate 1-carboxyvinyltransferase (EC 2.5.1.19)
- Lactococcus lactis >gi|683583|emb|CAA55180.1| (X78413)
5-enolpyruvylshikimate-3-phosphate synthase [Lactococcus
lactis]
/codon_start=1
/transl_table=1
/product="putative 3-phosphoshikimate
1-carboxyvinyltransferase"
/protein_id="AAK34180.1"
/db_xref="GI:13622461"
/translation="MKRMKLRTNAGPLQGTIQVPGDKSISHRVILGAVAKGETRVKG
LLKGEQVSTIOAFRLGVRIEKKDDQLVIEGOGFGLNAPCOTLNNSGTSMRLIA
GLLAGOPFSVKMGIDESLSKRPMDRTIVPLKMGVETSGETDRQFPLOLQGNRLQ
ITYTLPISSAOVKSAILLAAQAKGTQVVEREIRNHTHEMIQFEGRLIVDGKRIT
LVGPQGLTAGEITPDGDISAAPFWLVLVAGLIPGSELLKLVNPGVPTGILEVVERMG
AQIVEDMKKEQVTSIRVYVSNMKTIIISGLIPRLIDELPIALLATQAQGTTCIK
DAQELRVKTRQIVVTDILNSMGANIKATADGMIKGPVLYGANTSTYGDHRIGMM
TAAALLVKOGVHLDKEAINTSYPTFFKDLRLCHD"
complement(2013..2978)
/gene="SpY1353"
complement(2013..2978)
/notes="SpY1353"
/notes="Best Blastp hit = dbj|BAB07074.1| (AP001518)
unknown conserved protein [Bacillus halodurans]"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AAK34181.1"
/db_xref="GI:13622462"
/translation="MAEKWFKVLKWOYEIQVFMHLOSAMDLSAIVAYLIL
TAPPLVIAANFPYLINIDALLRLMKQLPKDIFRPAIAVENIFSPSGVLGVA
TLTGLTWGRSLTSLQAKINAYGASQHRDFFIGHVLGLTSLIIFLLAFALIFSI
SKAAIQVLKDHLSNITITFILLIQTIVLIVFLVGLMLLFLFLLPNVAKIKRIYILP
GTLTFSVMTGLFSLNLYGVVYVVERVDMKFGSVNFIIMLWFLFLARILILGAI
NATYQMSLGLKLEGRSGDMAILKTLGNSDLSPSQIEDSHTD"
complement(2980..3840)
/gene="map"
complement(2980..3840)
/notes="SpY1354"
complement(2980..3840)
/gene="map"
/EC_number="3.4.11.18"
/notes="Best Blastp hit = sp|Q92934|AMPM_BACHD METHIONINE
AMINOPEPTIDASE (MAP) (PEPTIDASE M)
>gi|11264635|pir||744405 methionyl aminopeptidase (EC
3.4.11.18) map [imported] - Bacillus halodurans
>gi|45124261|dbj|BAA75293.1| (AB017508) map homologue
(identity of 81% to B. subtilis ) [Bacillus halodurans]
>gi|10172768|dbj|BAB03875.1| (AP001507) methionine
aminopeptidase [Bacillus halodurans]"
/codon_start=1
/transl_table=1
/product="putative methionine aminopeptidase"
/protein_id="AAK34182.1"
/db_xref="GI:13622463"
/translation="MITLKSAREIEAMDRAQDFLAGIHIGLRLDIIRPGVDMWEVAYV
RRKREDNVLPQIGVDGHMMDYPATCCGLNDEVAHAPRHVILKEGDLKLVDMVLS
EPDKSIVDAALDFPNVPMKRWGTSGYTGGLADSCWAVGTPSDIEKOLMDVKEA
MYRGIEKAVIGNRIGDIGAQAQVYAESFGYGVYRDLVGHGCVGTPMHEEPVNYGTAG
RGLRLKEGMVLVTEPMINTGTWEIDTIDTGNWAKHTLDGSLSCQYEHQFVITKDPV
LTSQGEERT"
complement(3856..5139)
/gene="SpY1355"
complement(3856..5139)
/notes="SpY1355"
/notes="Best Blastp hit = pir||A69998 hypothetical protein
ytoi - Bacillus subtilis >gi|2293258|gb|AAC00336.1|

```

Alignment Scores:

```

(AF008220) ytoi [Bacillus subtilis]
>gi|2635332|emb|CAB14887.1| (Z99118) ytoi [Bacillus
subtilis]"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK34183.1"
/db_xref="GI:13622464"
/translation="MSKHQDILYLEKLAIGKVSYSRISNHLKVSQCTAYRAIKEAE
NRQIVETKSGTVRIEKKGRVIRDLTYSEIARISDVSEVLGAGLGHFESRPSIGA
MQONTERRVLVKGGLITVGDRETIQLLALEHNHAILVTGGFPVSKRVIEANNORIPV
MTHYDTFTVATMINHALSNIRKTDOLKTVQVMIPITDYGCESSVSEFNTLIK
TROVRFVLDYKRVKVGVSMDVQDPTTKLTKVMSKNPTARPNTSLANISOKMI
FEDLNLPTVDEENLLGLMITRQAMENLPHQPNPYTVSEILSNLETVDYQVYV
VEPTMDSAGNSMNGVISEFLKEISIRALKKQKNLIIQEMVYFLHAIQIEDELA
YPKLITENRNSMTDIEIFVDDQVIAKAIITTKIN"
complement(5148..5690)
/gene="SpY1356"
complement(5148..5690)
/notes="SpY1356"
/notes="Best Blastp hit = emb|CAA03913.1| (AJ000084)
putative acetyl transferase [Proteus mirabilis]"
/codon_start=1
/transl_table=11
/product="putative acetyl transferase"
/protein_id="AAK34184.1"
/db_xref="GI:13622465"
/translation="MDIWTKLAVFAFFETPKVILRPFRYEDHWDYFVSMVNDTNLYYV
FPEQTKAASDYLVLHSFIKPLGMAIEDKATHOVIGSIRIEHYDAKTRCADIYVFL
NVAFWGGQIMTEVYIKVLVLSFHEFGLKTLRIITHLENKASQKAVAKKQKLCFKG
SDRNTHKICIKMYOLTNR"
complement(5928..6581)
/gene="grab"
complement(5928..6581)
/notes="SpY1357"
complement(5928..6581)
/gene="grab"
/notes="Best Blastp hit = gb|AAD26340.1|AF124401.1
(AF124401) GRAB precursor [Streptococcus pyogenes]"
/codon_start=1
/transl_table=11
/product="protein GRAB (protein G-related alpha 2M-binding
protein)"
/protein_id="AAK34185.1"
/db_xref="GI:13622466"
/translation="MGKEIKVKCFLLRSAPGLVAVSASVLVSTVSAVDSPICPRII
PNGTTLNLGNAPKALNEERAIDELKQAIEDKEATTAIEAASSDALEALADQT
DALQSEEAAYVKADNAASDALEALADQTDALQSEAEVQSDNAASDAWKAATPIAL
DVKKTKDTRPVVKKEERQNVNTLPTTTEESNPFTAAALAIMYSTGLVYVSSCKEN"
complement(6939..8198)
/gene="murZ"
complement(6939..8198)
/notes="SpY1358"
complement(6939..8198)
/gene="murZ"
/notes="Best Blastp hit = sp|P19670|MUA2_BACSU
UDP-N-ACETYLGLUCOSAMINE 1-CARBOXYVINYLTRANSFERASE 2
(ENOYLPYRUVATE TRANSFERASE) (UDP-N-ACETYLGLUCOSAMINE
ENOYLPYRUVATE TRANSFERASE) (EPT) >gi|2127262|pir||G32354
UDP-N-acetylglucosamine 1-carboxyvinyltransferase (EC
2.5.1.7) murZ - Bacillus subtilis
>gi|853767|emb|CAA89875.1| (Z49782)
UDP-N-acetylglucosamine 1-carboxyvinyltransferase
[Bacillus subtilis] >gi|2636235|emb|CAB15727.1| (Z99122)
UDP-N-acetylglucosamine 1-carboxyvinyltransferase
[Bacillus subtilis] >gi|2636247|emb|CAB15738.1| (Z99123)
UDP-N-acetylglucosamine 1-carboxyvinyltransferase
[Bacillus subtilis]"
/codon_start=1
/transl_table=11
/product="putative UDP-N-acetylglucosamine
1-carboxyvinyltransferase"
/protein_id="AAK34186.1"

```



```

ORGANISM Streptococcus pyogenes
Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
Streptococcus.
REFERENCE 1 (bases 1 to 832)
AUTHORS Rasmussen, M., Muller, H.P. and Bjorck, L.
TITLE Protein GRAB of Streptococcus pyogenes regulates proteolysis at the
bacterial surface by binding alpha2-macroglobulin
JOURNAL J. Biol. Chem. 274 (22), 15336-15344 (1999)
MEDLINE 99269061
PUBMED 10336419
REFERENCE 2 (bases 1 to 832)
AUTHORS Rasmussen, M., Muller, H.P. and Bjorck, L.
TITLE Direct Submission
JOURNAL Submitted (28-JAN-1999) Cell and Molecular Biology, Molecular
Pathogenesis, Soelvegatan 39, Lund 221 00, Sweden
FEATURES
source
1. .832
/organism="Streptococcus pyogenes"
/strain="ATCC700294"
/db_xref="ATCC:700294"
/db_xref="taxon:1314"
1. .99
1. .654
/feature="grab"
/codon_start=1
/transl_table=11
/product="GRAB precursor"
/protein_id="AAD26338.1"
/db_xref="GI:4589079"
/translacion="MGKEIKVCFRLRSAGLVAASVLGVSTVSAVDSPIDPRII
PNGTLNLLAGPEKALNRERADIDELKQATIEDKEATTAIEAASDLEALADDT
DALQSEAAVKAADALADQTDALQSEAEVQSDNAAADAWKAATPIAL
DVKKTKDTPVYKKEQRQNTLTPTTGESNPFTAAALIMVSTGLVTVSSKCKEN"
100. .651
mat_peptide
100. .651
misc_feature
100. .273
/feature="GRAB"
/region: alpha2-macroglobulin-binding A domain"
274. .441
misc_feature
/region: repeat motif"
442. .576
misc_feature
/region: cell-wall attachment"
577. .651
misc_feature
/region: membrane-spanning"
304 a 143 c 169 g 216 t
BASE COUNT 304 a 143 c 169 g 216 t
ORIGIN
Alignment Scores:
Pred. No.: 1.04e-51 Length: 832
Score: 777.00 Matches: 159
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 1 Gaps: 0
US-09-847-539a-6 (1-159) x AF124399 (1-832)
Qy 1 ValAspSerProIleGluGlnProArgIleLeuProAsnGlyGlyThrLeuThrAsnLeu 20
|||||
Db 100 GTTGACTACCTATCGAACAGCTCGAATTTATCCAAATGGCGGAACCTTAACATACTT 159
Qy 21 LeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaIleAspGluLeu 40
|||||
Db 160 CTTTGCAATGCTCCAGAAAACCTGCATTAATGAAGAAGACCATTAATGAATTA 219
Qy 41 LysLysGlnAlaIleGluAspLysGluAlaThrAlaIleGluAlaAspSerAsp 60
|||||
Db 220 AAAAAACAGCTATTGAGATATAAGAGCTACGACAGCTATAGAAGCAGCAAGTTTCAGAT 279
Qy 61 AlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaValVal 80
|||||
Db 280 GCCTTAGAAGCATTAGCGGATCAACACAGCGCTTTACATCAGAGAAGCTCGGCTGCT 339
Qy 81 LysAlaAspAsnAlaAlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeu 100

```

```

|||||
Db 340 AAAGCGGATAACGCTGCTAGTGACGCTTAGAAGCATTTGGCGGATCAACAGACGCTTTA 399
Qy 101 GlnSerGluGluAlaGluValValGlnSerAspAsnAlaAlaSerAspAlaTrpGluLys 120
|||||
Db 400 CAATCAGAAGAGCTGAAGTAGTCTCAATCAGATAACGCTGCTAGTGACGCTGGGAAAAA 459
Qy 121 AlaAlaThrProIleAlaLeuAspValLysLysThrLysAspThrLysProValValLys 140
|||||
Db 460 GCAGCACTCCAAATGCTTTAGATGTTAAGAAAACATAAGATACAAAACCTGTAGTTAAA 519
Qy 141 LysGluGluArgGlnAsnValAsnThrLeuProThrGlyGluGluSerAsnPro 159
|||||
Db 520 AAGAAGAAAGACAAACAGTTAATACCTCTCTACAACTGTTGAAGAGTCTAACCCCA 576
RESULT 2
LOCUS AE006573 10029 bp DNA linear BCT 01-JUN-2001
DEFINITION Streptococcus pyogenes M1 GAS strain SF370, section 102 of 167 of
the complete genome.
ACCESSION AE006573 AE004092
VERSION AE006573.1 GI:13622459
KEYWORDS
SOURCE Streptococcus pyogenes M1 GAS.
ORGANISM Streptococcus pyogenes M1 GAS
Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
REFERENCE 1 (bases 1 to 10029)
AUTHORS Ferretti, J.J., McShan, W.M., Adjic, D., Savic, D., Savic, G., Lyon, K.,
Primeaux, C., Sezate, S.S., Surorov, A.N., Kenton, S., Lai, H., Lin, S.,
Qian, Y., Jia, H.G., Najjar, F.Z., Ren, Q., Zhu, H., Song, L., White, J.,
Yuan, X., Clifton, S.W., Roe, B.A. and McLaughlin, R.E.
Complete genome sequence of an M1 strain of Streptococcus pyogenes
Proc. Natl. Acad. Sci. U.S.A. 98 (8), 4658-4663 (2001)
JOURNAL 21192684
MEDLINE 11296296
PUBMED 11296296
REFERENCE 2 (bases 1 to 10029)
AUTHORS Ferretti, J.J., McShan, W.M., Adjic, D., Savic, D., Savic, G., Lyon, K.,
Primeaux, C., Sezate, S.S., Surorov, A.N., Kenton, S., Lai, H., Lin, S.,
Qian, Y., Jia, H.G., Najjar, F.Z., Ren, Q., Zhu, H., Song, L., White, J.,
Yuan, X., Clifton, S.W., Roe, B.A. and McLaughlin, R.E.
Direct Submission
Submitted (10-APR-2001) Department of Microbiology and Immunology,
University of Oklahoma Health Sciences Center, 940 SL Young Blvd,
Oklahoma City, OK 73104, USA
FEATURES
source
1. 10029
/organism="Streptococcus pyogenes M1 GAS"
/strain="SF370"
/serotype="M1"
/db_xref="taxon:160490"
complement(135. .626)
/feature="arok"
/feature="SPY1351"
complement(135. .626)
/feature="arok"
/EC_number="2.7.1.71"
/feature="Best Blastp hit = pir1ID82693 shikimate kinase
XF1335 [imported] - Xylella fastidiosa (strain 9a5c)
>gil9106332|gb|AF84144.1|AE003966_5 (AE003966) shikimate
kinase [Xylella fastidiosa]"
/codon_start=1
/transl_table=11
/product="putative shikimate kinase"
/protein_id="AAK34179.1"
/db_xref="GI:13622460"
/translacion="MTKVLGFMVGKTTVSKHLSMCKMDALIEAKIGMSIAAFPE
OHGFIAPRTIESVOLKLLFANDNSIIVTGGVYVQLQENRQLLRKHNNLLVASPE
TLYQLRKHDKKSQRPLFLKYSKEAFYEQRMVFYEGLSDLIVRDHRTPEEVANII
EGY"
complement(619. .1911)
/feature="arok"
/feature="SPY1352"

```

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 13, 2002, 04:49:23 ; Search time 1561.09 Seconds  
(without alignments)  
2131.407 Million cell updates/sec

Title: US-09-847-539A-6

Perfect score: 777

Sequence: 1 VDSPIEQRIIPNGGTLTNL.....KKEERQNVNTLPTTGEENP 159

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh  
-Q=/cgn2\_1/uspro\_spo01/JUS09847539/runat\_10102002\_093105\_4992/app\_query.fasta\_1.526  
-DB=GenEmbl -FAST=fastap -SUFFIX=std.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosu62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR\_SCORE=0.5 -THR\_MAX=100 -THR\_MIN=0 -ALIGN=40 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09847539@cgn\_1\_1\_1182@runat\_10102002\_093105\_4992 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.\*

1: gb\_ba.\*  
2: gb\_hgt.\*  
3: gb\_in.\*  
4: gb\_om.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_scs.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vi.\*  
15: em\_ba.\*  
16: em\_fun.\*  
17: em\_hum.\*  
18: em\_in.\*  
19: em\_mu.\*  
20: em\_on.\*  
21: em\_or.\*  
22: em\_ov.\*  
23: em\_pat.\*  
24: em\_ph.\*  
25: em\_pl.\*  
26: em\_ro.\*  
27: em\_scs.\*  
28: em\_un.\*

29: em\_vi.\*  
30: em\_hgt\_hum.\*  
31: em\_hgt\_inv.\*  
32: em\_hgt\_other.\*  
33: em\_hgtgo\_inv.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	777	100.0	832	1	AF124399	AF124399 Streptococ
2	777	100.0	10029	1	AE006573	AE006573 Streptoco
3	760	97.8	468	1	AF124401	AF124401 Streptoco
4	754	97.0	469	1	AF124400	AF124400 Streptoco
5	694	89.3	804	1	AF124403	AF124403 Streptoco
6	692	89.1	717	1	AF124402	AF124402 Streptoco
7	188	24.2	1950	1	STRSPGIGP	M13825 Streptococc
8	188	24.2	1950	6	IO9115	IO9115 Sequence 23
9	187	24.1	2384	1	SGSPG	Y00428 Streptococc
10	187	24.1	2384	6	IO9537	IO9537 Sequence 4
11	187	24.1	2456	1	SGPROTG	X06173 Streptococc
12	186	23.9	1950	6	IO9536	IO9536 Sequence 2
13	154.5	19.9	1576	1	SGI48PG	X5324 Streptococc
14	151.5	19.5	1380	1	SSCMLB1	X84989 streptococc
15	139	17.9	798	6	IO9107	IO9107 Sequence 4
16	138.5	17.8	969	1	STRABP	M95520 Streptococc
17	131.5	16.9	3838	3	PFASA7	M10330 Plasmodium
18	128.5	16.5	13501	1	AE005258	AE005258 Escherich
19	128.5	16.5	297816	1	AP002553	AP002553 Escherich
20	122.5	15.8	3763	5	PMU19361	U19361 Petromyzon
21	119.5	15.4	9331	1	AE006458	AE006458 Escherich
22	119.5	15.4	10432	1	AE005333	AE005333 Escherich
23	119.5	15.4	222605	1	AP002555	AP002555 Escherich
24	119.5	15.4	291136	1	AP002556	AP002556 Escherich
25	119.5	15.3	3222	14	HFU23857	U23857 Herpesvirus
26	117	15.1	54126	3	AC006834	AC006834 Caenorhab
27	116.5	15.0	146759	10	AL604063	AL604063 Mouse DNA
28	116	14.9	35318	3	CELT03A1	AF067614 Caenorhab
29	116	14.9	44090	2	AC006784	AC006784 Caenorhab
30	116	14.9	96468	2	AC006902	AC006902 Caenorhab
31	114.5	14.7	1555	1	STRMAG	L27798 Streptococc
32	114.5	14.7	106516	1	STYPPHGM2	AL513384 Salmonell
33	114.5	14.7	210791	2	AC023354	AC023354 Mus muscu
34	114.5	14.7	222304	2	AC104834	AC104834 Mus muscu
35	114	14.7	167227	9	AC020602	AC020602 Homo sapi
36	113	14.5	146491	2	AC019950	AC019950 Drosophil
37	112.5	14.5	4469	1	SPU52008	U52008 Streptococc
38	112	14.4	1469	6	SL148IGG	X04015 Streptococc
39	112	14.4	1469	6	AL12446	AL12446 1.5 kb Ecor
40	112	14.4	2025	5	AB025967	AB025967 Oryzias 1
41	111	14.3	1713	1	SPFCRA	X69324 S.pyogenes
42	111	14.3	2022	3	AY060997	AY060997 Drosophil
43	111	14.3	3120	1	AF324061	AF324061 Streptoco
44	111	14.3	6413	1	AF067776	AF067776 Abiotroph
45	110.5	14.2	48502	7	LAMCG	J02459 Bacterioph

# ALIGNMENTS

RESULT 1

AF124399

LOCUS

DEFINITION

complete cds.

ACCESSION

VERSION

KEYWORDS

SOURCE

AF124399 832 bp DNA linear BCT 14-AUG-2000  
Streptococcus pyogenes strain ATCC700294 GRAB precursor, gene.  
AF124399  
AF124399.1 GI:4589078  
Streptococcus pyogenes.

THIS PAGE BLANK (USPTO)

Job time : 338.257 secs



DB:		9	Gaps:	0
US-09-847-539A-6_COPY_59_86 (1-28) x AI062753 (1-666)				
Qy	2	AspaLaLeuGluAlaLeuAlaAspGlnThrAspaLaLeuGlnSerGluGluAlaAlaVal 21    :::             :::		
Db	227	GATCGCGGGAACTCATATAGGAACACACACGCCAGGAGGCAGGAGGAACAAGCCGCG 168		
Qy	22	VallysAlaAspAsnAla 27    :::		
Db	167	GTACGAGGAGATCGTGCC 150		
RESULT 36				
AZ699109/c				
LOCUS				
DEFINITION	AZ699109	666 bp DNA linear GSS 24-JAN-2001		
	RPCI-23-224J19.TV RPCI-23 Mus musculus genomic clone RPCI-23-224J19			
ACCESSION	AZ699109	DNA sequence.		
VERSION	AZ699109			
KEYWORDS	AZ699109.1 GI:12416865			
SOURCE	GSS.			
ORGANISM	house mouse.			
	Mus musculus			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 666) Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Aklinret .B., Levins,M., McGann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P., and Fraser,C.M. Mouse BAC End Sequences from Library RPCI-23			
TITLE	Unpublished (1999)			
JOURNAL	Other_GSSs: RPCI-23-224J19.TJ			
COMMENT	Contact: Shaying Zhao Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 0200 Fax: 301 838 0208 Email: szhao@tigr.org Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pjejong@email.chg.org). Clones may be purchased from BACPAC Resources ( <a href="http://www.chori.org/bacpac/orderingframe.htm">http://www.chori.org/bacpac/orderingframe.htm</a> ). BAC end page: <a href="http://ww.tigr.org/todb/bac_ends/mouse/bac_end_intro.html">http://ww.tigr.org/todb/bac_ends/mouse/bac_end_intro.html</a> Plate: 224 row: J column: 19 Seq primer: T7 Class: BAC ends.			

```

Class: BAC ends.
Location/Qualifiers
1. .666
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPC1-23-224J19"
/clone_lib="RPC1-23"
/sex="Female"
/lab_host="DH10B"
/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
ECORI; Site_2: EORI; Female C57BL/6J mouse kidney and/or
brain genomic DNA was isolated and partially digested
with a combination of EORI and EORI Methylase. Size
selected DNA was cloned into the pBACe3.6 vector at the
ECORI sites. The ligation products were transformed into
DH10B electrocompetent cells (BRL Life Technologies)."

BASE COUNT      203 a  163 c  150 g  150 t
ORIGIN

Alignment Scores:
Pred. No.:      172      Length:      666
Score:          51.00    Matches:      11
Percent Similarity: 68.18%  Conservat:      4
Best Local Similarity: 50.00%  Mismatches:      7
Query Match:      93.53%  Indels:      0
DB:              12      Gaps:      0

```

```

US-09-847-539A-6_COPY_59_86 (1-28) x AZ699109 (1-666)

Qy      1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAla 20
|||||  |||||  |||  ::||| ||||| ::||| ::|
Db      522 TCAGATCCTCGGAGACATTAAGAAATCAGACAGATTCCCTAAAGGCAGACAGAGAGAG 463
|||||  |||||  |||  ::||| ||||| ::||| ::|

Qy      21 ValVal 22
|||||  |||||
Db      462 GTAGTG 457

.RESULT 37
BI637204/c
LOCUS
DEFINITION
SD19151.5prime SD Drosophila melanogaster Schneider L2 cell culture
pOT2 Drosophila melanogaster cDNA clone SP19151 5 similar to
CG7289: FBAn0007289 GO:[] located on: 2L 22B8-22B8:: 05/19/2001,
mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Drosophila melanogaster
fruit fly.
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
AUTHORS
Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
Lewis,S. and Rubin,G.M.
TITLE
BGGP/HHMI Drosophila EST Project
JOURNAL
Unpublished (2001)
COMMENT
Contact: Stapleton, M.
BGGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST\_est@fruitfly.berkeley.edu
hit genomic AE003584: arm:2L [1824960,2149443]
estimated-cyto:22B4-22D2: 05/19/2001
Plate: SD.191 row: E column: 3
High quality sequence stop: 578.
Location/Qualifiers
1. 710
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="SD19151"
/clone_lib="SD Drosophila melanogaster Schneider L2 cell
culture pOT2"
/lab_host="DH5-alpha"
/notes="Vector: pOT2; Site_1: EcoRI; Site_2: XhoI; Sized
fractionated cDNAs were directly ligated into pOT2.
Plasmid cDNA library."
BASE COUNT 165 a 203 c 170 g 171 t 1 others
ORIGIN

Alignment Scores:
Pred. No.: 188 Length: 710
Score: 51.00 Matches: 11
Percent Similarity: 53.85% Conservative: 3
Best Local Similarity: 42.31% Mismatches: 12
Query Match: 39.53% Indels: 0
DB: 10 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x BI637204 (1-710)

Qy      2 AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaVal 21
|||||  |||  ::|  |||  |||||  |||  |||||  |||
Db      228 GATCGCGGAATCATATAGAAGACCAACACAGCCAGGAGGAGGAGCAACAGCGCG 169
|||||  |||  ::|  |||  |||||  |||  |||||  |||

Qy      22 VallysAlaAspAsnAla 27
|||||  |||  |||
Db      168 GTACGAGGAGATCGTGCC 151

.RESULT 38

```

```

Oy '22 Vallys 23
Db 5703 CTGGG 5708

RESULT 18
US-08-508-004-3
; Sequence 3, Application US/08508004
; Patent No. 5582969
; GENERAL INFORMATION:
; APPLICANT: Pearson, Robert E.
; APPLICANT: Dickson, Julie A.
; APPLICANT: Hamilton, Paul T.
; APPLICANT: Little, Michael C.
; APPLICANT: Beyrer Jr., Wayne F.
; TITLE OF INVENTION: MYCOBACTERIOPHAGE SPECIFIC FOR THE
; NUMBER OF INVENTION: MYCOBACTERIUM TUBERCULOSIS COMPLEX
; CORRESPONDENCE ADDRESSES: 6
; ADDRESSEE: Richard J. Rodrick, Becton Dickinson and
; ADDRESSEE: Company
; STREET: 1 Becton Drive
; CITY: Franklin Lakes
; STATE: NJ
; COUNTRY: US
; ZIP: 07417
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/508,004
; FILING DATE: 27-JUL-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/402,282
; FILING DATE: 10-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Fugit, Donna R.
; REGISTRATION NUMBER: 32,135
; REFERENCE/DOCKET NUMBER: P-3283
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15664 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 222..425
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 451..747
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 747..1109
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1109..2014
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2034..2747
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2747..3109
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 3109..3444
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 3444..3728
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 3731..4855
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 4855..5376
; OTHER INFORMATION: /function= "potential coding
; OTHER INFORMATION: sequence"
; OTHER INFORMATION: /product= "L5 gp37 homolog"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 5382..5747
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 5837..6307
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 6403..7770
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 7770..8006
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 8033..8236
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 8244..9443
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 9450..10244
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 10371..10586
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 11115..11786
; OTHER INFORMATION: /function= "potential open reading
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc_feature

```

LOCATION: 11917..12741  
OTHER INFORMATION: /function= "potential open reading"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 12748..14499  
OTHER INFORMATION: /function= "potential open reading"  
OTHER INFORMATION: frame"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 14771..15154  
OTHER INFORMATION: /function= "potential open reading"  
OTHER INFORMATION: frame"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 15154..15426  
OTHER INFORMATION: /function= "potential open reading"  
OTHER INFORMATION: frame"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 15429..15664  
OTHER INFORMATION: /function= "potential open reading"  
OTHER INFORMATION: frame"

US-08-508-004-3

Alignment Scores:  
Pred. No.: 2,66e+03 Length: 15664  
Score: 42.00 Matches: 8  
Percent Similarity: 63.64% Conservative: 6  
Best Local Similarity: 36.36% Mismatches: 8  
Query Match: 32.56% Indels: 0  
DB: 1 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-08-508-004-3 (1-15664)

QY 2 AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaVal 21  
::: ||| ::| ||||: ||| |||||  
Db 5643 GAGGGCTGCCCAATGACGACACAGCTCGAGTGTTCGACGACAGACGCGCGG 5702  
QY 22 ValLys 23  
::: ||||:  
Db 5703 CTGCGG 5708

RESULT 19  
US-08-402-066-3  
Sequence 3, Application US/08402066  
Patent No. 5612182  
GENERAL INFORMATION:  
APPLICANT: Pearson, Robert E.  
APPLICANT: Dickinson, Julie A.  
APPLICANT: Hamilton, Paul T.  
APPLICANT: Little, Michael C.  
APPLICANT: Beyer Jr., Wayne F.  
TITLE OF INVENTION: MYCOBACTERIOPHAGE SPECIFIC FOR THE  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Richard J. Rodrick, Becton Dickinson and  
ADDRESSEE: Company  
STREET: 1 Becton Drive  
CITY: Franklin Lakes  
STATE: NJ  
COUNTRY: US  
ZIP: 07417  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/402,066  
FILING DATE:  
CLASSIFICATION: 436

ATTORNEY/AGENT INFORMATION:  
NAME: Fugit, Donna R.  
REGISTRATION NUMBER: 32,135  
REFERENCE/DOCKET NUMBER: P-3283  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 15664 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 222..425  
OTHER INFORMATION: /function= "potential open reading"  
OTHER INFORMATION: frame"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 451..747  
OTHER INFORMATION: /function= "potential open reading"  
OTHER INFORMATION: frame"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 747..1109  
OTHER INFORMATION: /function= "potential open reading"  
OTHER INFORMATION: frame"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 1109..2014  
OTHER INFORMATION: /function= "potential open reading"  
OTHER INFORMATION: frame"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 2034..2747  
OTHER INFORMATION: /function= "potential open reading"  
OTHER INFORMATION: frame"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 2747..3109  
OTHER INFORMATION: /function= "potential open reading"  
OTHER INFORMATION: frame"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 3109..3444  
OTHER INFORMATION: /function= "potential open reading"  
OTHER INFORMATION: frame"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 3444..3728  
OTHER INFORMATION: /function= "potential open reading"  
OTHER INFORMATION: frame"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 3731..4855  
OTHER INFORMATION: /function= "potential open reading"  
OTHER INFORMATION: frame"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 4855..5376  
OTHER INFORMATION: /function= "potential coding"  
OTHER INFORMATION: sequence"  
OTHER INFORMATION: /product= "L5 gp37 homolog"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 5382..5747  
OTHER INFORMATION: /function= "potential open reading"  
OTHER INFORMATION: frame"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 5837..6307  
OTHER INFORMATION: /function= "potential open reading"  
OTHER INFORMATION: frame"  
FEATURE:



	/	NAME/KEY:	misc_feature			"potential open reading"
	/	LOCATION:	6403..7770			
	/	OTHER INFORMATION:	/function=			
	/	FEATURE:	FRAME			
	/	NAME/KEY:	misc_feature			
	/	LOCATION:	7770..8006			
	/	OTHER INFORMATION:	/function=			"potential open reading"
	/	FEATURE:	FRAME			
	/	NAME/KEY:	misc_feature			
	/	LOCATION:	8033-8236			
	/	OTHER INFORMATION:	/function=			"potential open reading"
	/	FEATURE:	FRAME			
	/	NAME/KEY:	misc_feature			
	/	LOCATION:	8244..9443			
	/	OTHER INFORMATION:	/function=			"potential open reading"
	/	FEATURE:	FRAME			
	/	NAME/KEY:	misc_feature			
	/	LOCATION:	9450..10244			
	/	OTHER INFORMATION:	/function=			"potential open reading"
	/	FEATURE:	FRAME			
	/	NAME/KEY:	misc_feature			
	/	LOCATION:	10371..10586			
	/	OTHER INFORMATION:	/function=			"potential open reading"
	/	FEATURE:	FRAME			
	/	NAME/KEY:	misc_feature			
	/	LOCATION:	11115..11786			
	/	OTHER INFORMATION:	/function=			"potential open reading"
	/	FEATURE:	FRAME			
	/	NAME/KEY:	misc_feature			
	/	LOCATION:	11917..12741			
	/	OTHER INFORMATION:	/function=			"potential open reading"
	/	FEATURE:	FRAME			
	/	NAME/KEY:	misc_feature			
	/	LOCATION:	12748..14499			
	/	OTHER INFORMATION:	/function=			"potential open reading"
	/	FEATURE:	FRAME			
	/	NAME/KEY:	misc_feature			
	/	LOCATION:	14771..15154			
	/	OTHER INFORMATION:	/function=			"potential open reading"
	/	FEATURE:	FRAME			
	/	NAME/KEY:	misc_feature			
	/	LOCATION:	15154..15426			
	/	OTHER INFORMATION:	/function=			"potential open reading"
	/	FEATURE:	FRAME			
	/	NAME/KEY:	misc_feature			
	/	LOCATION:	15429..15664			
	/	OTHER INFORMATION:	/function=			"potential open reading"
	/	FEATURE:	FRAME			
	/	US -08-402-066-3				
	/	Alignment Scores:				
	/	Pred. No.:	2.66e+03		Length:	15664
	/	Score:	42.00		Matches:	8
	/	Percent Similarity:	63.64%		Conservative:	6
	/	Best Local Similarity:	36.36%		Mismatches:	8
	/	Query Match:	32.56%		Indels:	0
	/	DB:	1		Gaps:	0
	/	US-09-847-539A-6_COPY_59_86 (1-28) x US-08-402-066-3 (1-15664)				
Oy		2 AspAlaLeuGluAlaLeuAspGlnThrAspAlaLeuGlnSerGluGlu				
	:	:::	::::	:	::::	:
	:	:::	::::	:	::::	:
	:	:::	::::	:	::::	:

```
; NAME/KEY: misc_feature
; LOCATION: 2747..3109
; OTHER INFORMATION: /function= "potential open reading"
; OTHER INFORMATION: frame=
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 3109..3444
; OTHER INFORMATION: /function= "potential open reading"
; OTHER INFORMATION: frame=
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 3444..3728
; OTHER INFORMATION: /function= "potential open reading"
; OTHER INFORMATION: frame=
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 3731..4855
; OTHER INFORMATION: /function= "potential open reading"
; OTHER INFORMATION: frame=
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 4855..5376
; OTHER INFORMATION: /function= "potential coding"
; OTHER INFORMATION: sequence=
; OTHER INFORMATION: /product= "L5 gp37 homolog"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 5382..5747
; OTHER INFORMATION: /function= "potential open reading"
; OTHER INFORMATION: frame=
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 5837..6307
; OTHER INFORMATION: /function= "potential open reading"
; OTHER INFORMATION: frame=
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 6403..7770
; OTHER INFORMATION: /function= "potential open reading"
; OTHER INFORMATION: frame=
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 7770..8006
; OTHER INFORMATION: /function= "potential open reading"
; OTHER INFORMATION: frame=
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 8033..8236
; OTHER INFORMATION: /function= "potential open reading"
; OTHER INFORMATION: frame=
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 8244..9443
; OTHER INFORMATION: /function= "potential open reading"
; OTHER INFORMATION: frame=
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 9450..10244
; OTHER INFORMATION: /function= "potential open reading"
; OTHER INFORMATION: frame=
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 10371..10586
; OTHER INFORMATION: /function= "potential open reading"
; OTHER INFORMATION: frame=
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 11115..11786
; OTHER INFORMATION: /function= "potential open reading"
; OTHER INFORMATION: frame=
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 11917..12741
```

```
; OTHER INFORMATION: /function= "potential open reading"
; OTHER INFORMATION: frame=
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 12748..14499
; OTHER INFORMATION: /function= "potential open reading"
; OTHER INFORMATION: frame=
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 14771..15154
; OTHER INFORMATION: /function= "potential open reading"
; OTHER INFORMATION: frame=
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 15154..15426
; OTHER INFORMATION: /function= "potential open reading"
; OTHER INFORMATION: frame=
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 15429..15664
; OTHER INFORMATION: /function= "potential open reading"
; OTHER INFORMATION: frame=
; US-08-402-068-3
Alignment Scores:
Pred. No.: 2.66e+03 Length: 15664
Score: 42.00 Matches: 8
Percent Similarity: 63.64% Conservative: 6
Best Local Similarity: 36.36% Mismatches: 8
Query Match: 32.56% Indels: 0
DB: 1 Gaps: 0
US-09-847-539A-6_COPY_59_86 (1-28) x US-08-402-068-3 (1-15664)
Qy 2 AspaLaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluAlaAlaVal 21
Db 5643 GAGGCGCTGGCCAACTGCAGACGCTGCTGTTGCAAGCAGACACGCGCGCG 5702
Qy 22 Vallys 23
Db 5703 CTGCGG 5708
RESULT 21
US-09-335-409-1
; Sequence 1, Application US/09335409
; Patent No. 6121029
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/335.409
; CURRENT FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-335-409-1
Alignment Scores:
Pred. No.: 1.65e+04 Length: 68750
Score: 42.00 Matches: 10
Percent Similarity: 59.26% Conservative: 6
Best Local Similarity: 37.04% Mismatches: 11
Query Match: 32.56% Indels: 0
DB: 3 Gaps: 0
```

```
US-09-847-539A-6_COPY_59_86 (1-28) x US-09-335-409-1 (1-68750)
QY 2 AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaVal 21
Db 53823 GATGCCCTCCGTCATGCCCGGAGGTCAGATCGTGGAGGCCGACGTG 53882

QY 22 VallysAlaAspAsnAlaAla 28
Db 53883 GCTCGGCGGACGATGTCGCT 53903

RESULT 22
US-09-568-102-1
; Sequence 1, Application US/09568102
; Patent No. 6346404
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,102
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-102-1

Alignment Scores:
Pred. No.: 1.65e+04 Length: 68750
Score: 42.00 Matches: 10
Percent Similarity: 59.26% Conservative: 6
Best Local Similarity: 37.04% Mismatches: 11
Query Match: 32.56% Indels: 0
DB: 4 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x US-09-567-969-1 (1-68750)
QY 2 AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaVal 21
Db 53823 GATGCCCTCCGTCATGCCCGGAGGTCAGATCGTGGAGGCCGACGTG 53882

QY 22 VallysAlaAspAsnAlaAla 28
Db 53883 GCTCGGCGGACGATGTCGCT 53903

RESULT 24
US-09-568-480-1
; Sequence 1, Application US/09568480
; Patent No. 6355458
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,480
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-480-1

Alignment Scores:
Pred. No.: 1.65e+04 Length: 68750
Score: 42.00 Matches: 10
Percent Similarity: 59.26% Conservative: 6
Best Local Similarity: 37.04% Mismatches: 11
Query Match: 32.56% Indels: 0
DB: 4 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x US-09-568-480-1 (1-68750)
QY 2 AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaVal 21
Db 53823 GATGCCCTCCGTCATGCCCGGAGGTCAGATCGTGGAGGCCGACGTG 53882

QY 22 VallysAlaAspAsnAlaAla 28
Db 53883 GCTCGGCGGACGATGTCGCT 53903

RESULT 25
US-09-568-486-1
; Sequence 1, Application US/09568486
; Patent No. 6355459
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
```



```
; Sequence 8, Application US/07876280
; Patent No. 5262158
; GENERAL INFORMATION:
; APPLICANT: Payne, Jewel M.
; APPLICANT: Cannon, Raymond J.C.
; APPLICANT: Bagley, Angela L.
; TITLE OF INVENTION: No. 5262158el Bacillus thuringiensis Isolates for
; TITLE OF INVENTION: Controlling Acarides
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David R. Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/876,280
; FILING DATE: 19920430
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Saliwanchik, David R.
; REGISTRATION NUMBER: 31,794
; REFERENCE/DOCKET NUMBER: M/S 104
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1185 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: BACILLUS THURINGIENSIS
; INDIVIDUAL ISOLATE: PS69D1
; IMMEDIATE SOURCE:
; CLONE: E. COLI NM522(pMYC2317) NRRL B-18816
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 1..1185
;
US-07-876-280-8
;
Alignment Scores:
Pred. No.: 163 Length: 1185
Score: 41.00 Matches: 9
Percent Similarity: 59.09% Conservative: 4
Best Local Similarity: 40.91% Mismatches: 9
Query Match: 31.78% Indels: 0
DB: 1 Gaps: 0
;
US-09-847-539a-6_COPY_59_86 (1-28) x US-07-876-280-8 (1-1185)
QY 5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaValVallysAla 24
Db 691 GAAGACTTACTAAGAAAGACTAGACGACCTGAAAAAAGAGAGTTGGAGCTGCTATTAAAGCA 750
QY 25 AspAsn 26
Db 751 GAGAAAT 756
;
RESULT 29
US-08-049-783-5
; Sequence 5, Application US/08049783
; Patent No. 5439881
;
GENERAL INFORMATION:
; APPLICANT: Narva, Kenneth E
; APPLICANT: Schwab, George E
; APPLICANT: Payne, Jewel M
; TITLE OF INVENTION: Gene Encoding No. 5439881el Nematode-Active
; TITLE OF INVENTION: Toxins Cloned from Bacillus thuringiensis Isolates
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Jeff Lloyd
; STREET: 2421 N.W. 41st Street
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/049,783
; FILING DATE: 19930419
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Lloyd, Jeff
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 5 (PS69D1):
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1185 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: BACILLUS THURINGIENSIS
; INDIVIDUAL ISOLATE: PS69D1
; IMMEDIATE SOURCE:
; LIBRARY: LAMBDA GEM(TW)-11 LIBRARY OF KENNETH NARVA
; CLONE: PS69D1A
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 1..1185
;
US-08-049-783-5
;
Alignment Scores:
Pred. No.: 163 Length: 1185
Score: 41.00 Matches: 9
Percent Similarity: 59.09% Conservative: 4
Best Local Similarity: 40.91% Mismatches: 9
Query Match: 31.78% Indels: 0
DB: 1 Gaps: 0
;
US-09-847-539a-6_COPY_59_86 (1-28) x US-08-049-783-5 (1-1185)
QY 5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaValVallysAla 24
Db 691 GAAGACTTACTAAGAAAGACTAGACGACCTGAAAAAAGAGAGTTGGAGCTGCTATTAAAGCA 750
QY 25 AspAsn 26
Db 751 GAGAAAT 756
;
RESULT 30
US-08-316-301A-9
; Sequence 9, Application US/08316301A
; Patent No. 5753492
; GENERAL INFORMATION:
; APPLICANT: Schnepf, Harry E.
; APPLICANT: Schwab, George E.
```

APPLICANT: Payne, Jewel M.  
APPLICANT: Narva, Kenneth E.  
APPLICANT: Fonceerrada, Luis  
TITLE OF INVENTION: No. 5753492el Nematode-Active Toxins and Genes  
TITLE OF INVENTION: Which Code Therefor  
NUMBER OF SEQUENCES: 42  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Saliwanchik & Saliwanchik  
STREET: 2421 N.W. 41st Street, Suite A-1  
CITY: Gainesville  
STATE: FL  
COUNTRY: USA  
ZIP: 32606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/316.301A  
FILING DATE: 30-SEP-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/871,510  
FILING DATE: 23-APR-1992  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/693,018  
FILING DATE: 03-MAY-1991  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/565,544  
FILING DATE: 10-AUG-1990  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/084,653  
FILING DATE: 12-AUG-1987  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/830,050  
FILING DATE: 31-JAN-1992  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Lloyd, Jeff  
REGISTRATION NUMBER: 35,589  
REFERENCE/DOCKET NUMBER: MA20CCCD1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 904-375-8100  
TELEFAX: 904-372-5800  
INFORMATION FOR SEQ ID NO: 9 (PS69D1):  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1185 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: BACILLUS THURINGIENSIS  
INDIVIDUAL ISOLATE: PS69D1  
IMMEDIATE SOURCE:  
CLONE: E. coli NM522(pMYC2317) NRRL B-18816  
FEATURE:  
NAME/KEY: mat\_peptide  
LOCATION: 1..1185  
US-08-316-301A-9  
Alignment Scores:  
Pred. No.: 163  
Score: 41.00  
Percent Similarity: 59.09%  
Best Local Similarity: 40.91%  
Length: 1185  
Matches: 9  
Conservative: 4  
Mismatch: 9

Query Match: 31.78% Indels: 0  
DB: 1 Gaps: 0  
US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-08-316-301A-9 (1-1185)  
Qy 5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVallyAla 24  
Db 691 CAAGAGTTTACTAAAGAAAGTAGACGACCTGAAAGAGAGTTTGAAGAGCTGCTATTAAAGCA 750  
Qy 25 AspAsn 26  
Db 751 GAGAA 756  
RESULT 31  
US-08-904-278-9  
Sequence 9, Application US/08904278  
Patent No. 5874288  
GENERAL INFORMATION:  
APPLICANT: Thompson, Mark  
APPLICANT: Knuth, Mark  
APPLICANT: Cardineau, Guy  
TITLE OF INVENTION: Bacillus thuringiensis Toxins With Improved  
TITLE OF INVENTION: Activity  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Saliwanchik & Saliwanchik  
STREET: 2421 N.W. 41st Street, Suite A-1  
CITY: Gainesville  
STATE: Florida  
COUNTRY: USA  
ZIP: 32606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/904,278  
FILING DATE:  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Sanders, Jay M.  
REGISTRATION NUMBER: 39,355  
REFERENCE/DOCKET NUMBER: MA-702  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (352) 375-8100  
TELEFAX: (352) 372-5800  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1185 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: BACILLUS THURINGIENSIS  
INDIVIDUAL ISOLATE: PS69D1  
IMMEDIATE SOURCE:  
CLONE: E. coli NM522(pMYC2317) NRRL B-18816  
FEATURE:  
NAME/KEY: mat\_peptide  
LOCATION: 1..1185  
US-08-904-278-9  
Alignment Scores:  
Pred. No.: 163  
Score: 41.00  
Percent Similarity: 59.09%  
Best Local Similarity: 40.91%  
Length: 1185  
Matches: 9  
Conservative: 4  
Mismatch: 9  
DB: 2 Indels: 0









## TELECOMMUNICATION INFORMATION:

TELEPHONE: 312-715-1000

TELEFAX: 312-715-1234

TELEX: 910-221-5317

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 1680 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

FEATURE:

NAME/KEY: 5'UTR

LOCATION: 1..30

FEATURE:

NAME/KEY: CDS

LOCATION: 31..1626

FEATURE:

NAME/KEY: 3'UTR

LOCATION: 1626..1680

US-08-916-745-2

Alignment Scores:

Pred. No.: 252

Score: 1680

Matches: 10

Conservative: 3

Best Local Similarity: 56.52%

Mismatches: 10

Query Match: 31.78%

Indels: 0

DB: 0

US-09-847-539a-6\_COPY\_59\_86 (1-28) x US-08-916-745-2 (1-1680)

QY 5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLysAla 24

Db 399 GAGGCCAAAGTAGCGGACGCGATGCCCGACAGAGCCCGAGGAGCGCTGGCATCGAGCGA 340

QY 25 AspAsnAla 27

Db 339 GCGCGCGCC 331

RESULT 39

US-09-042-929-2/c

Sequence 2, Application US/09042929

Patent No. 5919628

GENERAL INFORMATION:

APPLICANT: Amara, Susan G

APPLICANT: Arriza, Jeffrey L

TITLE OF INVENTION: Amino Acid Transporters and Uses

NUMBER OF SEQUENCES: 17

CORRESPONDENCE ADDRESS:

ADDRESS: Allegretti &amp; Witcoff, Ltd.

STREET: 10 South Wacker Drive, Suite 3000

CITY: Chicago

STATE: IL

COUNTRY: USA

ZIP: 60606

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/042,929

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/140,729

FILING DATE: 20 OCT 1993

ATTORNEY/AGENT INFORMATION:

NAME: No. 5919628nan, Kevin E

REGISTRATION NUMBER: 35,303

REFERENCE/DOCKET NUMBER: 93,509

TELECOMMUNICATION INFORMATION:

TELEPHONE: 312-715-1000

TELEFAX: 312-715-1234

TELEX: 910-221-5317

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 1680 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

FEATURE:

NAME/KEY: 5'UTR

LOCATION: 1..30

FEATURE:

NAME/KEY: CDS

LOCATION: 31..1626

FEATURE:

NAME/KEY: 3'UTR

LOCATION: 1626..1680

US-09-042-929-2

Alignment Scores:

Pred. No.: 252

Score: 1680

Matches: 10

Conservative: 3

Best Local Similarity: 56.52%

Mismatches: 10

Query Match: 31.78%

Indels: 0

DB: 0

US-09-847-539a-6\_COPY\_59\_86 (1-28) x US-09-042-929-2 (1-1680)

QY 5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLysAla 24

Db 399 GAGGCCAAAGTAGCGGACGCGATGCCCGACAGAGCCCGAGGAGCGCTGGCATCGAGCGA 340

QY 25 AspAsnAla 27

Db 339 GCGCGCGCC 331

RESULT 40

US-08-546-661-2/c

Sequence 2, Application US/08546661

Patent No. 5919699

GENERAL INFORMATION:

APPLICANT: Amara, Susan G

APPLICANT: Arriza, Jeffrey L

TITLE OF INVENTION: Amino Acid Transporters and Uses

NUMBER OF SEQUENCES: 17

CORRESPONDENCE ADDRESS:

ADDRESS: Allegretti &amp; Witcoff, Ltd.

STREET: 10 South Wacker Drive, Suite 3000

CITY: Chicago

STATE: IL

COUNTRY: USA

ZIP: 60606

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/546,661

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/140,729

FILING DATE: 20 OCT 1993

ATTORNEY/AGENT INFORMATION:

NAME: No. 5919699nan, Kevin E

REGISTRATION NUMBER: 35,303

REFERENCE/DOCKET NUMBER: 93,509

TELECOMMUNICATION INFORMATION:

TELEPHONE: 312-715-1000



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: October 13, 2002, 00:48:06 ; Search time 11.3797 seconds  
(without alignments)  
273.300 Million cell updates/sec

Title: US-09-847-539A-6\_COPY\_59\_86

Perfect score: 129

Sequence: 1 SDALALADQTDALQSEAAVVRADNAA 28

Scoring table: BLOSUM62

Gapop 60.0 , Gapext 60.0

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

## Database :

A\_Geneseq\_032802.\*  
1: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	129	100.0	28	AA1980	Streptococcus pyog
2	129	100.0	141	AA1981	Streptococcus pyog
3	129	100.0	155	AA1982	Streptococcus pyog
4	129	100.0	159	AA1983	Streptococcus pyog
5	129	100.0	167	AA1984	Streptococcus pyog
6	129	100.0	184	AA1985	Streptococcus pyog
7	129	100.0	217	AA1986	Streptococcus pyog
8	129	100.0	259	AA1987	Streptococcus pyog
9	129	100.0	271	AA1988	Streptococcus pyog
10	57	44.2	192	AA1989	H. pylori secreted
11	57	44.2	222	AA1990	H. pylori secreted

12	57	44.2	412	19	AA10994	H. pylori ORF 09c
13	50	38.8	1164	22	ABG30268	Novel human diago
14	47	36.4	258	20	AA136771	Chlamydia trachoma
15	46	35.7	705	22	AA190301	C glutamicum prote
16	45	34.9	502	21	AA154580	A rat calcium/cal
17	45	34.9	1362	20	AA199481	Murine p/cip prote
18	45	34.9	1447	20	AA181029	Murine p/cip prote
19	45	34.9	1637	22	ABG60342	Drosophila melanog
20	44	34.1	42	22	ABG01077	Novel human diago
21	44	34.1	274	21	AA158293	Arabidopsis thalia
22	44	34.1	274	21	AA181028	Arabidopsis thalia
23	44	34.1	284	22	AA178512	Human protein seq
24	44	34.1	296	21	AA158292	Arabidopsis thalia
25	44	34.1	296	21	AA181027	Arabidopsis thalia
26	44	34.1	344	19	AA161234	Streptococcus pneu
27	44	34.1	358	21	AA158291	Arabidopsis thalia
28	44	34.1	358	21	AA181026	Arabidopsis thalia
29	44	34.1	371	21	AA181672	Streptococcus pneu
30	44	34.1	421	22	AB171817	Drosophila melanog
31	44	34.1	429	22	AB165976	Drosophila melanog
32	44	34.1	514	19	AA180400	A secreted protein
33	44	34.1	592	22	AA192957	C glutamicum prote
34	44	34.1	951	20	AA181028	ER interacting dom
35	44	34.1	1412	22	AA159278	Human SRC-3 protei
36	44	34.1	1420	20	AA181025	Human SRC-3 protei
37	44	34.1	1522	20	AA121975	Human steroid rece
38	44	34.1	1630	22	AB159874	Drosophila melanog
39	43	33.3	33	21	AA195962	Yeast GCN4 mutated
40	43	33.3	92	21	AA14712	Shorthorn sculpin
41	43	33.3	226	21	AA121230	Tomato H-207 prote
42	43	33.3	254	22	AA12557	Chlamydia pneumoni
43	43	33.3	258	20	AA135145	Chlamydia pneumoni
44	43	33.3	407	20	AA105072	S. pneumoniae folC
45	43	33.3	418	20	AA105071	S. pneumoniae folC

## ALIGNMENTS

RESULT 1  
AA171038  
ID AA171038 standard; peptide: 28 AA.  
XX  
AC AA171038;  
XX  
DT 29-AUG-2000 (first entry)  
XX  
DE Streptococcus pyogenes strain SF370 GRAB protein fragment #3.  
XX  
KW GRAB protein; protein G related alpha2M binding protein; vaccine;  
KW alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;  
KW immune response; Streptococcus pyogenes infection.  
XX  
OS Streptococcus pyogenes.  
XX  
PN WO200026240-A2.  
XX  
PD 11-MAY-2000.  
XX  
PF 02-NOV-1999; 99WO-GB03631.  
XX  
PR 02-NOV-1998; 98GB-0023975.  
XX  
PA (ACTI-) ACTINOVA LTD.  
XX  
PI Bjorck LH, Rasmussen M;  
XX  
DR WPI; 2000-365572/31.  
XX  
PT New alpha2M binding protein for generating a protective immune response  
PT to group A streptococcus and purifying the binding protein -  
XX  
PS Claim 3; Page 56; 67pp; English.



```
RESULT 4
AAV71041
ID AAV71041 standard; peptide; 159 AA.
XX AC
XX AC AAV71041;
DT 29-AUG-2000 (first entry)
XX ST
DE Streptococcus pyogenes strain SF370 GRAB protein fragment #5.
XX GRAB protein; protein G related alpha2M binding protein; vaccine;
KW alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;
KW immune response; Streptococcus pyogenes infection.
XX OS
OS Streptococcus pyogenes.
XX PN
XX WO200026240-A2.
XX PD
XX PD 11-MAY-2000.
XX PF
XX PF 02-NOV-1999; 99WO-GB03631.
XX PR
XX PR 02-NOV-1998; 98GB-0023975.
XX PA
XX PA (ACTI-) ACTINOVA LTD.
XX PI
XX PI Bjorck LH, Rasmussen M;
XX WPI; 2000-365572/31.
XX DR
XX DR N-PSDB; AAD00564.
XX ST
XX ST New alpha2M binding protein for generating a protective immune response
to group A streptococcus and purifying the binding protein -
XX PS
XX PS Claim 5; Page 57-58; 67pp; English.
XX CC
XX CC The patent discloses a new family of proteins termed GRAB (protein G
related alpha2M binding protein) from Streptococcus pyogenes which have
the ability to bind alpha2-macroglobulin (alpha2M) and show homology to
protein G of group G Streptococcus. GRAB protein and peptides derived
from it are used in vaccine compositions for generating a protective
immune response against group A Streptococcus. Antibodies against GRAB
are useful for treating Streptococcus pyogenes infections. The protein
is also useful for purifying alpha2M from a sample. The present sequence
is a fragment of GRAB protein from S. pyogenes strain SF370 corresponding
to residues 34-192 and devoid of the membrane spanning region. This
fragment is useful in vaccine composition.
XX SQ
SQ Sequence 159 AA;
Query Match 100.0%; Score 129; DB 21; Length 159;
Best Local Similarity 100.0%; Pred. No. 3.4e-11;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 SDALALADQDTALQSEEAAYVVKADNAA 28
Db 59 SDALALADQDTALQSEEAAYVVKADNAA 86
RESULT 5
AAV71046
ID AAV71046 standard; Protein; 167 AA.
XX AC
XX AC AAV71046;
DT 29-AUG-2000 (first entry)
XX ST
DE Streptococcus pyogenes strain KTL3 partial GRAB protein.
XX GRAB protein; protein G related alpha2M binding protein; vaccine;
KW alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;
KW immune response; Streptococcus pyogenes infection.
XX SQ
```

```
OS Streptococcus pyogenes.
XX PN
XX WO200026240-A2.
XX PD
XX PD 11-MAY-2000.
XX PF
XX PF 02-NOV-1999; 99WO-GB03631.
XX PR
XX PR 02-NOV-1998; 98GB-0023975.
XX PA
XX PA (ACTI-) ACTINOVA LTD.
XX PI
XX PI Bjorck LH, Rasmussen M;
XX WPI; 2000-365572/31.
XX DR
XX DR N-PSDB; AAD00564.
XX ST
XX ST New alpha2M binding protein for generating a protective immune response
to group A streptococcus and purifying the binding protein -
XX PS
XX PS Claim 5; Page 62-63; 67pp; English.
XX CC
XX CC The patent discloses a new family of proteins termed GRAB (protein G
related alpha2M binding protein) from Streptococcus pyogenes which have
the ability to bind alpha2-macroglobulin (alpha2M) and show homology to
protein G of group G Streptococcus. GRAB protein and peptides derived
from it are used in vaccine compositions for generating a protective
immune response against group A Streptococcus. Antibodies against GRAB
are useful for treating Streptococcus pyogenes infections. The protein
is also useful for purifying alpha2M from a sample. The present sequence
is a partial GRAB protein from S. pyogenes strain KTL3.
XX CC
XX CC The protein has alpha2M binding region and is useful in vaccine
composition.
XX SQ
SQ Sequence 167 AA;
Query Match 100.0%; Score 129; DB 21; Length 167;
Best Local Similarity 100.0%; Pred. No. 3.6e-11;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 SDALALADQDTALQSEEAAYVVKADNAA 28
Db 70 SDALALADQDTALQSEEAAYVVKADNAA 97
RESULT 6
AAV71039
ID AAV71039 standard; Protein; 184 AA.
XX AC
XX AC AAV71039;
DT 29-AUG-2000 (first entry)
XX ST
DE Streptococcus pyogenes strain SF370 mature GRAB protein.
XX GRAB protein; protein G related alpha2M binding protein; vaccine;
KW alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;
KW immune response; Streptococcus pyogenes infection.
XX OS
OS Streptococcus pyogenes.
XX FH
XX FH Key Location/Qualifiers
FT Binding-site 1..58
FT Region /label= alpha2-macroglobulin_binding_site
FT Region 59..86
FT Region /label= R1
FT Region /note= "repeat region"
FT Region 87..114
FT Region /label= R2
FT Region /note= "repeat region"
FT Region 115..159
FT Region /label= Cell_wall_spanning_region
FT Region 150..155
```

```

FT FT /note= "consensus sequence for gram-positive
FT FT surface cell wall anchored proteins"
FT FT 160..184
FT FT /label= Membrane_spanning_region
XX
XX WO200026240-A2.
XX
XX 11-MAY-2000.
XX
XX 02-NOV-1999; 99WO-GB03631.
XX
XX 02-NOV-1998; 98GB-0023975.
XX
XX (ACTI-) ACTINOVA LTD.
XX
XX Bjorck LH, Rasmussen M;
XX
XX WPI; 2000-365572/31.
XX
XX New alpha2M binding protein for generating a protective immune response
XX to group A streptococcus and purifying the binding protein
XX
XX Claim 5; Page 56; 67pp; English.
XX
XX The patent discloses a new family of proteins termed GRAB (protein G
XX related alpha2M binding protein) from Streptococcus pyogenes which have
XX the ability to bind alpha2-macroglobulin (alpha2M) and show homology to
XX protein G of group G Streptococcus. GRAB protein and peptides derived
XX from it are used in vaccine compositions for generating a protective
XX immune response against group A Streptococcus. Antibodies against GRAB
XX are useful for treating Streptococcus pyogenes infections. The protein
XX is also useful for purifying alpha2M from a sample. The present sequence
XX is a mature GRAB protein from S. pyogenes strain SF370 without the
XX signal sequence. This sequence is capable of binding alpha2M
XX and useful in vaccine composition.
XX
XX SQ Sequence 184 AA;
XX
XX Query Match 100.0%; Score 129; DB 21; Length 184;
XX Best Local Similarity 100.0%; Pred. No. 4e-11;
XX Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 SDALEALADQTDALQSEEAAYVKADNAA 28
XX |
XX 59 SDALEALADQTDALQSEEAAYVKADNAA 86
XX
XX RESULT 7
XX AAY71042
XX ID AAY71042 standard; Protein; 217 AA.
XX
XX AC AAY71042;
XX
XX DT 29-AUG-2000 (first entry)
XX
XX STREPTOCOCCUS PYOGENES STRAIN SF370 FULL-LENGTH GRAB PROTEIN.
XX
XX GRAB protein; protein G related alpha2M binding protein; vaccine;
XX alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;
XX immune response; Streptococcus pyogenes infection.
XX
XX STREPTOCOCCUS PYOGENES.
XX
XX OS Location/Qualifiers
XX FH 1..33
XX FT /label= Signal_sequence
XX FT 34..217
XX FT /label= Mature_GRAB_protein
XX FT Binding-site 34..91
XX FT /label= alpha2M_binding_site
XX FT Domain 34..68
XX FT /note= "shows homology to E domain of protein G"
XX FT Region 92..119

```

```

FT Region /label= Repeat_region_1
FT 120..147
FT /label= Repeat_region_2
FT 148..192
FT /label= Cell_wall_spanning_region
FT 183..188
FT /note= "consensus sequence for gram-positive
FT surface cell wall anchored proteins"
FT 193..217
FT /label= Membrane_spanning_region
XX
XX WO200026240-A2.
XX
XX 11-MAY-2000.
XX
XX 02-NOV-1999; 99WO-GB03631.
XX
XX 02-NOV-1998; 98GB-0023975.
XX
XX (ACTI-) ACTINOVA LTD.
XX
XX Bjorck LH, Rasmussen M;
XX
XX WPI; 2000-365572/31.
XX N-PSDB; AAD00559, AAD00560.
XX
XX New alpha2M binding protein for generating a protective immune response
XX to group A streptococcus and purifying the binding protein
XX
XX Claim 5; Fig 2B; 67pp; English.
XX
XX The patent discloses a new family of proteins termed GRAB (protein G
XX related alpha2M binding protein) from Streptococcus pyogenes which have
XX the ability to bind alpha2-macroglobulin (alpha2M) and show homology to
XX protein G of group G Streptococcus. GRAB protein and peptides derived
XX from it are used in vaccine compositions for generating a protective
XX immune response against group A Streptococcus. Antibodies against GRAB
XX are also useful for purifying alpha2M from a sample. The present sequence
XX is a full-length GRAB protein from S. pyogenes strain SF370.
XX
XX SQ Sequence 217 AA;
XX
XX Query Match 100.0%; Score 129; DB 21; Length 217;
XX Best Local Similarity 100.0%; Pred. No. 4.9e-11;
XX Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 SDALEALADQTDALQSEEAAYVKADNAA 28
XX |
XX 92 SDALEALADQTDALQSEEAAYVKADNAA 119
XX
XX RESULT 8
XX AAY71043
XX ID AAY71043 standard; Protein; 259 AA.
XX
XX AC AAY71043;
XX
XX DT 29-AUG-2000 (first entry)
XX
XX STREPTOCOCCUS PYOGENES STRAIN KTL9 PARTIAL GRAB PROTEIN.
XX
XX GRAB protein; protein G related alpha2M binding protein; vaccine;
XX alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;
XX immune response; Streptococcus pyogenes infection.
XX
XX STREPTOCOCCUS PYOGENES.
XX
XX OS WO200026240-A2.
XX FT 11-MAY-2000.
XX
XX 02-NOV-1999; 99WO-GB03631.

```



```

XX PR 02-NOV-1998; 98GB-0023975.
XX PA (ACTI-) ACTINOVA LTD.
XX PI Bjorck LH, Rasmussen M;
XX DR WPI; 2000-365572/31.
XX DR N-PSDB; AAD00561.
XX PT New alpha2M binding protein for generating a protective immune response
XX PT to group A streptococcus and purifying the binding protein -
XX PS Claim 5; Page 59-60; 67pp; English.
XX CC The patent discloses a new family of proteins termed GRAB (protein G
XX CC related alpha2M binding protein) from Streptococcus pyogenes which have
XX CC the ability to bind alpha2-macroglobulin (alpha2M) and show homology to
XX CC protein G of group G Streptococcus. GRAB protein and peptides derived
XX CC from it are used in vaccine compositions for generating a protective
XX CC immune response against group A Streptococcus. Antibodies against GRAB
XX CC are useful for treating Streptococcus pyogenes infections. The protein
XX CC is also useful for purifying alpha2M from a sample. The present sequence
XX CC is a partial GRAB protein from S. pyogenes strain AP49.
XX CC The protein has alpha2M binding region and is useful in vaccine
XX CC composition.
XX SQ Sequence 259 AA;
XX Query Match 100.0%; Score 129; DB 21; Length 259;
XX Best Local Similarity 100.0%; Pred. No. 6e-11;
XX Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 SDALEALADQTDALQSEEAHVVKADNAA 28
XX DB 107 SDALEALADQTDALQSEEAHVVKADNAA 134
XX
XX RESULT 9
XX AAY71045
XX ID AAY71045 standard; Protein; 271 AA.
XX AC AAY71045;
XX DT 29-AUG-2000 (first entry)
XX DE Streptococcus pyogenes strain AP49 partial GRAB protein.
XX KW GRAB protein; protein G related alpha2M binding protein; vaccine;
XX KW alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;
XX KW immune response; Streptococcus pyogenes infection.
XX OS Streptococcus pyogenes.
XX PN WO200026240-A2.
XX PD 11-MAY-2000.
XX PF 02-NOV-1999; 99WO-GB03631.
XX PR 02-NOV-1998; 98GB-0023975.
XX PA (ACTI-) ACTINOVA LTD.
XX PI Bjorck LH, Rasmussen M;
XX DR WPI; 2000-365572/31.
XX DR N-PSDB; AAD00563.
XX PT New alpha2M binding protein for generating a protective immune response
XX PT to group A streptococcus and purifying the binding protein -
XX PS Claim 5; Page 61-62; 67pp; English.
XX
XX CC The patent discloses a new family of proteins termed GRAB (protein G
XX CC related alpha2M binding protein) from Streptococcus pyogenes which have
XX CC the ability to bind alpha2-macroglobulin (alpha2M) and show homology to
XX CC protein G of group G Streptococcus. GRAB protein and peptides derived
XX CC from it are used in vaccine compositions for generating a protective
XX CC immune response against group A Streptococcus. Antibodies against GRAB
XX CC are useful for treating Streptococcus pyogenes infections. The protein
XX CC is also useful for purifying alpha2M from a sample. The present sequence
XX CC is a partial GRAB protein from S. pyogenes strain KTL9.
XX CC The protein has alpha2M binding region and is useful in vaccine
XX CC composition.
XX SQ Sequence 271 AA;
XX Query Match 100.0%; Score 129; DB 21; Length 271;
XX Best Local Similarity 100.0%; Pred. No. 6.3e-11;
XX Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 SDALEALADQTDALQSEEAHVVKADNAA 28
XX DB 174 SDALEALADQTDALQSEEAHVVKADNAA 201
XX
XX RESULT 10
XX AAW20445
XX ID AAW20445 standard; Protein; 192 AA.
XX AC AAW20445;
XX DT 11-JUL-1997 (first entry)
XX DE H. pylori secreted or periplasmic protein, 35336707.aa.
XX KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
XX KW identification; binding compound; bacterium; life cycle; activator;
XX KW bacteria; inhibitor; duodenal ulcer disease; chronic gastritis;
XX KW diagnosis.
XX OS Helicobacter pylori.
XX FH Key Location/Qualifiers
XX FT Misc-difference 181 /note= "encoded by GYG"
XX FT Misc-difference 183 /note= "encoded by GYC"
XX FT Misc-difference 187 /note= "encoded by AWT"
XX FT Misc-difference 192 /note= "encoded by AWT"
XX PN WO9640893-A1.
XX PD 19-DEC-1996.
XX PF 06-JUN-1996; 96WO-US09122.
XX PR 01-APR-1996; 96US-0630405.
XX PR 07-JUN-1995; 95US-0487032.
XX PA (ASTR ) ASTRA AB.
XX PI Berglindh OT, Smith D, Mellgaard BL;
XX DR WPI; 1997-052306/05.
XX DR N-PSDB; AAT67618.
XX PT Helicobacter pylori nucleic acid sequences and related
XX PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori
XX PT infection, and to detect Helicobacter
XX PS Claim 72; Page 619; 1481pp; English.

```

CC This sequence represents a *H. pylori* secreted or periplasmic protein.  
 CC The protein may be used in a vaccine to prevent or treat *H. pylori*  
 CC infection or to identify *H. pylori* polypeptide binding compounds,  
 CC useful as potential *H. pylori* life cycle activators or inhibitors.  
 CC The genomic sequence of *H. pylori* (ATCC 55679) was determined from  
 CC overlapping contigs generated by mechanically shearing the bacterial  
 CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,  
 CC and the predicted coding regions defined by computer evaluation. To  
 CC identify likely *H. pylori* antigens for vaccine development, the amino  
 CC acid sequences predicted from various ORF were analysed for significant  
 CC homology to other known or exported membrane proteins. Having identified  
 CC and determined the sequences of interest, particular regions can be  
 CC isolated from *H. pylori* by PCR amplification for recombinant polypeptide  
 CC production, e.g. in *E. coli* hosts.

XX SQ Sequence 192 AA;

Query Match 44.2%; Score 57; DB 18; Length 192;  
 Best Local Similarity 52.2%; Pred. No. 1.3;  
 Matches 12; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 4 LEALADOTDALQSEAAVVKADN 26

Db 49 IQALQEQIDALDSQEKVSKWDN 71

RESULT 11

AAW20638  
 ID AAW20638 standard; protein; 222 AA.

XX AC AAW20638;

XX DT 14-JUL-1997 (first entry)

XX DE *H. pylori* secreted or periplasmic protein, 02cel0216orf1.

XX KW Cytoplasmic; vaccine; prevention; treatment; infection; identification;  
 KW binding compound; bacterium; life cycle; activator; bacteria; inhibitor;  
 KW duodenal ulcer disease; chronic gastritis; diagnosis; envelope.  
 XX OS Helicobacter pylori.

XX FN WO9640893-AL.

XX PD 19-DEC-1996.

XX PF 06-JUN-1996; 96WO-US09122.

XX PR 01-APR-1996; 96US-0630405.

XX PR 07-JUN-1995; 95US-0487032.

XX PA (ASTR ) ASTRA AB.

XX PI Berglindh OT, Smith D, Mellgaerd BL;

XX DR WPI; 1997-052306/05.  
 DR N-PSDB; AAT67891.

XX PT Helicobacter pylori nucleic acid sequences and related  
 PT polypeptide(s) - useful for vaccines to treat or prevent *H. pylori*  
 PT infection, and to detect Helicobacter

XX PS Claim 72; Page 1062; 1481pp; English.

XX CC The present sequence is a *H. pylori* secreted or periplasmic protein.  
 CC The protein may be used in a vaccine to prevent or treat *H. pylori*  
 CC infection or to identify *H. pylori* polypeptide binding compounds,  
 CC useful as potential *H. pylori* life cycle activators or inhibitors.  
 CC The genomic sequence of *H. pylori* (ATCC 55679) was determined from  
 CC overlapping contigs generated by mechanically shearing the bacterial  
 CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,  
 CC and the predicted coding regions defined by computer evaluation. To  
 CC identify likely *H. pylori* antigens for vaccine development, the amino

CC acid sequences predicted from various ORF were analysed for significant  
 CC homology to other known or exported membrane proteins. Having identified  
 CC and determined the sequences of interest, particular regions can be  
 CC isolated from *H. pylori* by PCR amplification for recombinant polypeptide  
 CC production, e.g. in *E. coli* hosts.

XX SQ Sequence 222 AA;

Query Match 44.2%; Score 57; DB 18; Length 222;  
 Best Local Similarity 52.2%; Pred. No. 1.5;  
 Matches 12; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 4 LEALADOTDALQSEAAVVKADN 26

Db 77 IQALQEQIDALDSQEKVSKWDN 99

RESULT 12

AAV10994  
 ID AAV10994 standard; protein; 412 AA.

XX AC AAV10994;

XX DT 08-JUN-1999 (first entry)

XX DE *H. pylori* ORF 09cel0413\_35336707\_f2\_9 secreted protein.

XX KW Vaccine; probe; diagnostic; ORF; cell envelope protein;  
 KW secreted protein; cellular protein.

XX OS Helicobacter pylori.

XX FN WO9818323-AL.

XX PD 07-MAY-1998.

XX PF 28-OCT-1997; 97WO-US19575.

XX PR 14-JUL-1997; 97US-0891928.

XX PR 28-OCT-1996; 96US-0739150.

XX PR 06-DEC-1996; 96US-0759739.

XX PA (ASTR ) ASTRA AB.

XX PI Alm RA, Smith D;

XX DR WPI; 1998-271811/24.  
 DR N-PSDB; AAX30461.

XX PT Helicobacter pylori nucleic acids and proteins - used to develop  
 PT products for the detection, prevention and treatment of *H. pylori*  
 PT infections

XX PS Claims 27, 31; Page 201-202; 279pp; English.

XX CC Recombinant or substantially pure preparations of *H. pylori* polypeptides  
 CC are disclosed, together with the nucleic acids encoding them. In all,  
 CC 73 ORFs are shown. The proteins are variously cell envelope proteins,  
 CC secreted proteins or other cellular proteins. Vaccines containing the  
 CC nucleic acids or proteins are claimed, as are probes containing at least  
 CC 8 nucleotides from the nucleic acid sequences. The vaccines are useful  
 CC for treating or reducing the risk of *H. pylori* infections, and the  
 CC probes can be used diagnostically for detecting the presence of  
 CC Helicobacter in a sample. The products are also of use in screening  
 CC for compounds having the ability to interfere with the *H. pylori* life  
 CC cycle or to inhibit *H. pylori* infection.

XX SQ Sequence 412 AA;

Query Match 44.2%; Score 57; DB 19; Length 412;  
 Best Local Similarity 52.2%; Pred. No. 3.1;  
 Matches 12; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

```
QY 4 LEALADQTDALQSEEAAYVKADN 26
   ::|| :||| :||| :|||
Db 49 IQALQEQIDALDSQEKVSKWDN 71

RESULT 13
ID ABC30268 standard; Protein; 1164 AA.
XX
AC ABC30268;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #30259.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI: 2001-639362/73.
DR N-PSDB; AAS94455.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 20; SEQ ID No 60627; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABC00010-ABC30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1164 AA;
   Query Match 38.8%; Score 50; DB 22; Length 1164;
   Best Local Similarity 45.8%; Pred. No. 1.1e+02;
   Matches 11; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

QY 3 ALEALADQTDALQSEEAAYVKADN 26
   | : ||| :||| :|||
Db 49 IQALQEQIDALDSQEKVSKWDN 71

RESULT 14
ID AAY36771 standard; Protein; 258 AA.
XX
AC AAY36771;
XX
DT 07-OCT-1999 (first entry)
XX
DE Chlamydia trachomatis cellular envelope protein.
XX
KW Vaccine; eye disease; conventional trachoma; nonendemic trachoma;
KW paratrachoma; inclusion conjunctivitis; genital disease; peritritis;
KW nongonococcal urethritis; epididymitis; cervicitis; salpingitis;
KW bartholinitis; pneumopathy; venereal lymphogranulomatosis.
XX
OS Chlamydia trachomatis.
XX
PN WO9928475-A2.
XX
PD 10-JUN-1999.
XX
PF 27-NOV-1998; 98WO-IB01939.
XX
PR 04-NOV-1998; 98US-0107077.
PR 28-NOV-1997; 97FR-0015041.
PR 17-DEC-1997; 97FR-0016034.
XX
PA (GEST ) GENSET.
XX
PI Griffais R;
XX
DR WPI: 1999-371125/31.
XX
PT Genome sequence of Chlamydia trachomatis
PS Disclosure; Page 668-669; 1755pp; English.
XX
CC AAY36754-Y37949 are encoded by open reading frames (ORFs) of the genome
CC of Chlamydia trachomatis (see AA201425). The polypeptides can be used as
CC vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences
CC can also be used to control growth of the microorganism. Chlamydia
CC trachomatis is responsible for a large number of diseases, e.g. eye
CC diseases such as conventional trachoma, nonendemic trachoma,
CC paratrachoma, and inclusion conjunctivitis; genital diseases such as
CC nongonococcal urethritis, epididymitis, cervicitis, salpingitis,
CC peritritis, bartholinitis; pneumopathy in breast feeding infants;
CC and venereal lymphogranulomatosis. The polypeptides of the invention
CC may be of use in treating these diseases.
XX
SQ Sequence 258 AA;
   Query Match 36.4%; Score 47; DB 20; Length 258;
   Best Local Similarity 36.4%; Pred. No. 51;
   Matches 8; Conservative 9; Mismatches 5; Indels 0; Gaps 0;

QY 4 LEALADQTDALQSEEAAYVKAD 25
   ::||| :||| :||| :|||
Db 71 IQEISDQINKLENQQAAYVKMD 92

RESULT 15
ID AAG90301 standard; Protein; 705 AA.
XX
AC AAG90301;
XX
DT 26-SEP-2001 (first entry)
XX
DE C glutamicum protein fragment SEQ ID NO: 4055.
XX
```

KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;  
 KW organic acid synthesis.  
 XX  
 OS Corynebacterium glutamicum.  
 XX  
 XX EPI108790-A2.  
 PN  
 XX 20-JUN-2001.  
 PD  
 XX 18-DEC-2000; 2000EP-0127688.  
 XX  
 XX 16-DEC-1999; 99JP-0377484.  
 PR 07-APR-2000; 2000JP-0159162.  
 PR 03-AUG-2000; 2000JP-0280988.  
 XX  
 XX (KYOW ) KYOWA HAKKO KOGYO KK.  
 PA  
 XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;  
 PI Tateishi N, Senoh A, Ikeda M, Ozaki A;  
 XX  
 XX WPI; 2001-376931/40.  
 DR N-PSDB; AAH65520.  
 XX  
 XX Novel polynucleotides derived from Coryneform bacteria, for identifying  
 PT mutation point of a gene, measuring expression of a gene, analysing  
 PT expression profile or pattern of a gene and identifying homologous gene  
 PT  
 XX  
 PS Claim 17; SEQ ID NO: 4055; 246pp + Sequence Listing; English.  
 XX  
 CC The present invention provides a number of nucleotide and protein  
 CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These  
 CC are useful for identifying the mutation point of a gene derived from a  
 CC mutant of coryneform bacterium, measuring expression amount and  
 CC analyzing the expression profile or expression pattern of a gene derived  
 CC from Coryneform bacterium, and identifying a homologue of a gene derived  
 CC from coryneform bacterium. Coryneform bacteria are useful for producing  
 CC amino acids, nucleic acids, vitamins, saccharides and organic acids,  
 CC particularly L-lysine. The present sequence is a protein described  
 CC in the exemplification of the invention.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from the  
 CC European Patent Office.  
 XX  
 SQ Sequence 705 AA;  
 Query Match 35.78; Score 46; DB 22; Length 705;  
 Best Local Similarity 37.58; Pred. No. 2.3e+02;  
 Matches 9; Conservative 6; Mismatches 9; Indels 0; Gaps 0;  
 Qy 4 LEALADQTDALQSEEAAYVKADNA 27  
 Db 514 IEPYAPEADELEGESEAYKPEA 537  
 RESULT 16  
 AAY54580  
 ID AAY54580 standard; Protein; 502 AA.  
 XX  
 AC AAY54580;  
 XX  
 XX 25-APR-2000 (first entry)  
 DT  
 XX A rat calcium/calmodulin dependent protein kinase designated CaMK-IV.  
 DE  
 XX Calcium/calmodulin dependent protein kinase; CaMK-VI; CaMK-IV;  
 KW CaMK-VI related peptide; CARP; neuronal plasticity; epilepsy;  
 KW kainate-induced seizure; CARP; neuronal plasticity; epilepsy;  
 KW kainate-induced seizure; seizure; central nervous system; hippocampus;  
 KW acute brain insult; stroke; trauma; hypoxia; ischemia;  
 KW neurodegenerative disease; Alzheimer's disease.  
 XX  
 OS Rattus sp.  
 XX

FH Key Location/Qualifiers  
 FT Domain 62..324  
 FT /note= "putative catalytic domain"  
 FT Binding-site 99  
 FT /note= "residue involved in ATP binding"  
 FT Misc-difference 244  
 FT /note= "residue thought to be crucial for  
 FT activation of the kinase by phosphorylation"  
 FT Domain 227..234  
 FT /note= "Ser/Thr kinase motif"  
 FT Domain 346..364  
 FT /note= "regulatory domain"  
 XX EP978562-A1.  
 PN  
 XX 09-FEB-2000.  
 PD  
 XX 06-AUG-1998; 98EP-0202654.  
 XX  
 XX 06-AUG-1998; 98EP-0202654.  
 XX  
 XX (MEDT-) DIV MEDICAL PHARMACOLOGY LEIDEN AMSTERDA.  
 PA  
 XX WPI; 2000-138770/13.  
 DR  
 XX Mammalian calcium/calmodulin dependent protein kinase related protein  
 XX useful as a modulator of calcium/calmodulin dependent protein kinase  
 PT activity, useful for treating seizures and damage to the central  
 PT nervous system -  
 PT  
 XX Disclosure; Page 34-36; 48pp; English.  
 XX  
 CC The present sequence represents a calcium/calmodulin dependent protein  
 CC kinase (CaMK-IV). The specification also describes another CaMK,  
 CC designated CaMK-VI, and a CaMK-VI related peptide, designated CARP. The  
 CC CARP peptide is produced by alternative splicing of the CaMK-VI gene,  
 CC and plays a role in kinase-mediated neuronal plasticity which is  
 CC associated with epilepsy. A large part of CARP (48 out of 55 amino  
 CC acids) is highly homologous with the carboxy-terminus of human  
 CC doublecortin. It is possible the calcium overload in a cell, triggered  
 CC by kainate-induced seizures, leads to activation of CaMK-VI. This  
 CC enhanced kinase activity induces CARP expression, which subsequently  
 CC acts as a modulator of the kinase, e.g. by acting as a substrate or  
 CC by irreversible binding to the activated kinase. The mammalian CARP  
 CC protein may be used for the treatment and/or prevention of seizures  
 CC (e.g. in epilepsy) and/or damage to the central nervous system,  
 CC particularly the hippocampus. CARP peptides may also be used to treat  
 CC acute brain insults (e.g. resulting from stroke, trauma, hypoxia, and  
 CC ischemia), and neurodegenerative diseases (e.g. Alzheimer's disease).  
 XX  
 SQ Sequence 502 AA;  
 Query Match 34.98; Score 45; DB 21; Length 502;  
 Best Local Similarity 58.8%; Pred. No. 2.1e+02;  
 Matches 10; Conservative 3; Mismatches 4; Indels 0; Gaps 0;  
 Qy 2 DALEALADQTDALQSEEE 18  
 Db 406 EAKAAADETMKLOSEE 422  
 RESULT 17  
 AAW99481  
 ID AAW99481 standard; Protein; 1362 AA.  
 XX  
 AC AAW99481;  
 XX  
 XX 08-JUN-1999 (first entry)  
 DT  
 XX Murine p/CIP protein.  
 DE  
 XX Mouse; p/CIP; p300/CBP/co-integrator-associated protein; gene expression;  
 KW breast cancer; inflammatory disease; atherosclerosis; osteoporosis; ds.





PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 11-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 21-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.

PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 28-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.





```
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 11-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.

PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 34.1%; Score 44; DB 21; Length 274;
Best Local Similarity 39.1%; Pred. No. 1.5e+02;
Matches 9; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

QY 5 EALADQTDLQSEEAHVVKADNA 27
   ||| : || : |:: |||
Db 53 EALQQRDKALSERDKALIERDNA 75

RESULT 23
AAM78512
ID AAM78512 standard; Protein; 284 AA.
XX
AC AAM78512;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human protein SEQ ID NO 1174.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation.
XX
OS Homo sapiens.
XX
PN WO200157190-A2.
XX
PD
XX
PF 09-AUG-2001.
XX
PF 05-FEB-2001; 2001WO-US04098.
PR 03-FEB-2000; 2000US-0496914.
PR 27-APR-2000; 2000US-0560875.
PR 20-JUN-2000; 2000US-0598075.
PR 19-JUL-2000; 2000US-0620325.
PR 01-SEP-2000; 2000US-0654936.
PR 15-SEP-2000; 2000US-0663561.
PR 20-OCT-2000; 2000US-0693325.
PR 30-NOV-2000; 2000US-0728422.
XX
(PHYSE-) HYSEQ INC.
XX
Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX
WPI: 2001-476283/51.
DR N-PSDB; AAK51645.
XX
Nucleic acids encoding polypeptides with cytokine-like activities,
PT useful in diagnosis and gene therapy -
XX
Claim 20; Page 3417-3418; 6221pp; English.
PS
XX
The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAK80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
```

CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation.  
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666  
CC (AAM80020) are omitted as the relevant pages from the sequence listing  
CC were missing at the time of publication.

XX SQ Sequence 284 AA;  
Query Match 34.18; Score 44; DB 22; Length 284;  
Best Local Similarity 47.68; Pred. No. 1.6e+02;  
Matches 10; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

Oy 5 BALADQTDALQSEAAVVKAD 25  
Db 62 EALKDAEDSLLAEEAAKAE 82

RESULT 24  
AAG36292  
ID AAG58292 standard; Protein; 296 AA.

XX AC AAG58292;

DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 75230.

KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.

OS Arabidopsis thaliana.

XX EP1033405-A2.

XX PD 06-SEP-2000.

XX PF 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126284.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132407.  
PR 05-MAY-1999; 99US-0132484.  
PR 06-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.

99US-0135124.  
PR 20-MAY-1999; 99US-0135353.  
PR 21-MAY-1999; 99US-0135629.  
PR 24-MAY-1999; 99US-0136021.  
PR 25-MAY-1999; 99US-0136392.  
PR 27-MAY-1999; 99US-0136782.  
PR 28-MAY-1999; 99US-0137222.  
PR 01-JUN-1999; 99US-0137528.  
PR 03-JUN-1999; 99US-0137502.  
PR 04-JUN-1999; 99US-0137724.  
PR 07-JUN-1999; 99US-0138094.  
PR 08-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 28-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.

```
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157863.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.

PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 34.1%; Score 44; DB 21; Length 296;
Best Local Similarity 39.1%; Pred. No. 1.6e+02;
Matches 9; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

QY 5 EALADQTDALQSEEAAYVKADNA 27
DB 75 EALQQRDKALSERDKALIERDNA 97
||| : || : |||
75 EALQQRDKALSERDKALIERDNA 97

RESULT 25
AAG61027
ID AAG61027 standard; Protein; 296 AA.
XX
AC AAG61027;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 79108.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EPL033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
```

PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139452.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139753.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 28-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.

PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 26-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151086.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161922.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match

34.1%; Score 44; DB 21; Length 296;

Best Local Similarity 39.1%; Pred. No. 1.6e+02;  
Matches 9; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

QY 5 EALADQTDALQSEEAHVVDNA 27  
DB 75 EALQQRDKALSERDKALIERDNA 97

## RESULT 26

AAW61234  
ID AAW61234 standard; Protein; 344 AA.

XX AC AAW61234;

XX DT 02-OCT-1998 (first entry)

XX DE Streptococcus pneumoniae Sp109 protein.

XX KW Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;  
detection; pneumonia; otitis media; meningitis.

XX OS Streptococcus pneumoniae.

XX PN W09818930-A2.

XX PD 07-MAY-1998.

XX PF 30-OCT-1997; 97WO-US19422.

XX PR 31-OCT-1996; 96US-0029960.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Choi GH, Hromockyj A, Johnson LS, Kunsch CA;

XX DR WPI; 1998-272224/24.

XX DR N-PSDB; AAV27419.

XX PT Nucleic acid encoding antigenic peptide(s) from Streptococcus  
pneumoniae - or their epitope-containing fragments, useful in  
protective or therapeutic vaccines, and for diagnosis

XX PS Claim 11; Page 87; 118pp; English.

XX CC The present sequence represents a protein from Streptococcus pneumoniae.  
CC The nucleic acid sequence encoding the Streptococcus pneumoniae protein  
CC can be useful in vaccines for inducing protective antibodies against  
CC Streptococcus pneumoniae, for treatment or prevention of infection e.g.  
CC pneumonia, otitis media or meningitis. Probes based on the nucleic acid  
CC are used to detect Streptococcus infection (by usual hybridisation or  
CC amplification methods), also for isolating Streptococcus genes or their  
CC allelic variants. The protein can be used similarly to detect specific  
CC antibodies in standard immunoassays, especially for diagnosing or  
CC monitoring infections. Antibodies which bind the protein are used to  
CC detect corresponding antigens, to purify the protein and for passive  
CC immunisation (optionally coupled to a toxin). Vaccines are administered,  
CC e.g. by injection, orally or through the skin, typically at 0.01-1000  
CC (especially 10-300) mu g/ml per dose.

XX SQ Sequence 344 AA;

Query Match 34.1%; Score 44; DB 19; Length 344;

Best Local Similarity 55.0%; Pred. No. 1.9e+02;  
Matches 11; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 8 AQOTDQLQSEEAHVVDNA 27

DB 3 AGQTDAQIEKAASQGGKA 22

## RESULT 27

AAG58291  
ID AAG58291 standard; Protein; 358 AA.

XX AAG58291;  
XX DT 18-OCT-2000 (first entry)  
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 75229.  
XX KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
XX termination sequence.  
XX OS Arabidopsis thaliana.  
XX PN EP1033405-A2.  
XX PD 06-SEP-2000.  
XX PF 25-FEB-2000; 2000EP-0301439.  
XX PR 25-FEB-1999; 99US-0121825.  
XX PR 05-MAR-1999; 99US-0123180.  
XX PR 09-MAR-1999; 99US-0123548.  
XX PR 23-MAR-1999; 99US-0125788.  
XX PR 25-MAR-1999; 99US-0126264.  
XX PR 29-MAR-1999; 99US-0126785.  
XX PR 01-APR-1999; 99US-0127462.  
XX PR 06-APR-1999; 99US-0128234.  
XX PR 08-APR-1999; 99US-0128714.  
XX PR 16-APR-1999; 99US-0129845.  
XX PR 19-APR-1999; 99US-0130077.  
XX PR 21-APR-1999; 99US-0130449.  
XX PR 23-APR-1999; 99US-0130510.  
XX PR 28-APR-1999; 99US-0130891.  
XX PR 30-APR-1999; 99US-0131449.  
XX PR 30-APR-1999; 99US-0132048.  
XX PR 04-MAY-1999; 99US-0132407.  
XX PR 05-MAY-1999; 99US-0132485.  
XX PR 06-MAY-1999; 99US-0132486.  
XX PR 06-MAY-1999; 99US-0132487.  
XX PR 07-MAY-1999; 99US-0132863.  
XX PR 11-MAY-1999; 99US-0134256.  
XX PR 14-MAY-1999; 99US-0134218.  
XX PR 14-MAY-1999; 99US-0134219.  
XX PR 14-MAY-1999; 99US-0134221.  
XX PR 14-MAY-1999; 99US-0134370.  
XX PR 18-MAY-1999; 99US-0134768.  
XX PR 19-MAY-1999; 99US-0134941.  
XX PR 20-MAY-1999; 99US-0135124.  
XX PR 21-MAY-1999; 99US-0135353.  
XX PR 24-MAY-1999; 99US-0135629.  
XX PR 25-MAY-1999; 99US-0136021.  
XX PR 27-MAY-1999; 99US-0136392.  
XX PR 28-MAY-1999; 99US-0136782.  
XX PR 01-JUN-1999; 99US-0137222.  
XX PR 03-JUN-1999; 99US-0137528.  
XX PR 04-JUN-1999; 99US-0137502.  
XX PR 07-JUN-1999; 99US-0137724.  
XX PR 08-JUN-1999; 99US-0138094.  
XX PR 10-JUN-1999; 99US-0138540.  
XX PR 10-JUN-1999; 99US-0138847.  
XX PR 16-JUN-1999; 99US-0139119.  
XX PR 16-JUN-1999; 99US-0139452.  
XX PR 16-JUN-1999; 99US-0139453.  
XX PR 17-JUN-1999; 99US-0139492.  
XX PR 18-JUN-1999; 99US-0139454.  
XX PR 18-JUN-1999; 99US-0139455.  
XX PR 18-JUN-1999; 99US-0139456.  
XX PR 18-JUN-1999; 99US-0139457.  
XX PR 18-JUN-1999; 99US-0139458.  
XX PR 18-JUN-1999; 99US-0139459.  
XX PR 18-JUN-1999; 99US-0139460.  
XX PR 18-JUN-1999; 99US-0139461.

PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149369.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.

PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 34.1%; Score 44; DB 21; Length 358;  
Best Local Similarity 39.1%; Pred. No. 2e+02;  
Matches 9; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

QY 5 EALADQTDALQSEEAAYVKA DNA 27  
||| : || : ||: |||  
Db 137 EALQQRDKALSERDKALIERDNA 159

RESULT 28  
AAG61026  
ID AAG61026 standard; Protein; 358 AA.  
XX  
AC AAG61026;  
XX  
DT 18-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 79107.  
XX

KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
XX termination sequence.

OS Arabidopsis thaliana.

PN EP1033405-A2.

XX

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

XX

25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 28-APR-1999; 99US-0130891.

PR 28-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 30-APR-1999; 99US-0132407.

PR 04-MAY-1999; 99US-0132484.

PR 05-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

PR 07-MAY-1999; 99US-0132487.

PR 07-MAY-1999; 99US-0132863.

PR 11-MAY-1999; 99US-0134256.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134221.

PR 14-MAY-1999; 99US-0134370.

PR 18-MAY-1999; 99US-0134768.

PR 19-MAY-1999; 99US-0134941.

PR 20-MAY-1999; 99US-0135124.

PR 21-MAY-1999; 99US-0135353.

PR 24-MAY-1999; 99US-0135629.

PR 25-MAY-1999; 99US-0136021.

PR 27-MAY-1999; 99US-0136392.

PR 28-MAY-1999; 99US-0136782.

PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 21-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 05-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.





DR WPI; 2001-656860/75.  
DR N-PSDB; ABL15920.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX  
PS Disclosure; SEQ ID NO 42243; 21pp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins  
CC (ABB57737-ABB72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 421 AA;  
  
Query Match 34.1%; Score 44; DB 22; Length 421;  
Best Local Similarity 46.2%; Pred. No. 2.4e+02;  
Matches 12; Conservative 3; Mismatches 11; Indels 0; Gaps 0;  
  
QY 3 ALEALADQTDALQSEEAAYVKADNAA 28  
| | | | | : | | | | | : | |  
Db 348 AACAPADEAAPAAPEEAAPPPAEAA 373  
  
RESULT 31  
ABB65976  
ID ABB65976 standard; Protein; 429 AA.  
XX  
AC ABB65976;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster polypeptide SEQ ID NO 24720.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical.  
XX  
OS Drosophila melanogaster.  
XX  
PN WO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US09231.  
XX  
PR 23-MAR-2000; 2000US-191637P.  
PR 11-JUL-2000; 2000US-0614150.  
XX  
PA (PEKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li PWD, Myers EW;  
XX  
DR WPI; 2001-656860/75.  
DR N-PSDB; ABL10079.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX  
PS Disclosure; SEQ ID NO 24720; 21pp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins  
CC (ABB57737-ABB72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 429 AA;  
  
Query Match 34.1%; Score 44; DB 22; Length 429;  
Best Local Similarity 48.1%; Pred. No. 2.5e+02;  
Matches 13; Conservative 2; Mismatches 12; Indels 0; Gaps 0;  
  
QY 2 DALEALADQTDALQSEEAAYVKADNAA 28  
: | | | | : | | | | | : | | | | |  
Db 182 NAAEVLAAAGNADESVAAPAAQNAA 208  
  
RESULT 32  
AAW80400  
ID AAW80400 standard; Protein; 514 AA.  
XX  
AC AAW80400;  
XX  
DT 12-JAN-1999 (first entry)  
XX  
DE A secreted protein encoded by clone dd71\_2.  
XX  
KW Secreted protein; immune stimulating; suppressing;  
KW haematopoiesis regulating activity; tissue growth activity; activin;  
KW inhibin activity; chemotactic; chemokinetic activity; haemostatic;  
KW thrombolytic activity; anti-inflammatory activity; cadherin;  
KW tumour invasion suppressor activity; tumour inhibition activity.  
XX  
OS Homo sapiens.  
XX  
PN WO9840486-A2.  
XX  
PD 17-SEP-1998.  
XX  
PF 13-MAR-1998; 98WO-US04977.  
XX  
PR 29-OCT-1997; 97US-0960022.  
PR 14-MAR-1997; 97US-0815047.  
XX  
PA (GEMY ) GENETICS INST INC.  
XX  
PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;  
PI Racie LA, Spaulding V, Treacy M;  
XX  
DR WPI; 1998-520812/44.  
DR N-PSDB; AAV63174.  
XX  
PT New isolated human poly:nucleotide(s) and secreted proteins -  
PT obtained from e.g. human foetal kidney, placenta, foetal brain,  
PT adult testes, adult brain or adult uterus cDNA libraries  
XX  
PS Claim 29; Pages 80-82; 110pp; English.  
XX  
CC The present sequence represents a secreted protein. The nucleic acid  
CC sequence is isolated from a human adult testes cDNA library using  
CC probe AAV63184. The polypeptide may have biological activities such as  
CC e.g. nutritional activity, immune stimulating or suppressing activity,  
CC haematopoiesis regulating activity, tissue growth activity,  
CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory  
CC activity, cadherin/tumour invasion suppressor activity, tumour  
CC inhibition activity or other activities.  
XX  
SQ Sequence 514 AA;  
  
Query Match 34.1%; Score 44; DB 19; Length 514;

Best Local Similarity 36.0%; Pred. No. 3.1e+02;  
Matches 9; Conservative 8; Mismatches 8; Indels 0; Gaps 0;

OY 3 ALEALADQTDALQSEEAAYVVKADNA 27  
DB 240 SIAQRQAQTOLLKEEARLVESNNA 264  
:::| | | :||| :||| :|||

## RESULT 33

AAG92957  
ID AAG92957 standard; Protein: 592 AA.

XX AC AAG92957;

DT 26-SEP-2001 (first entry)

XX DE C glutamincum protein fragment SEQ ID NO: 6711.

XX KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;

XX OS organic acid synthesis.

XX PN Corynebacterium glutamicum.

XX PD EP1108790-A2.

XX PF 20-JUN-2001.

XX PR 18-DEC-2000; 2000EP-0127688.

XX PR 16-DEC-1999; 99JP-0377484.

XX PR 07-APR-2000; 2000JP-0159162.

XX PR 03-AUG-2000; 2000JP-0280988.

XX PA (KYOW ) KYOWA HAKKO KOGYO KK.

XX PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochial K, Yokoi H;

XX PI Tateishi N, Senoh A, Ikeda M, Ozaki A;

XX DR WPI; 2001-376931/40.

XX DR N-PSDB; AAH68176.

XX PT Novel polynucleotides derived from Coryneform bacteria, for identifying

XX PT mutation point of a gene, measuring expression of a gene, analysing

XX PT expression profile or pattern of a gene and identifying homologous gene

XX PS Claim 17; SEQ ID NO: 6711; 246pp + Sequence Listing: English.

XX CC The present invention provides a number of nucleotide and protein

XX CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These

XX CC are useful for identifying the mutation point of a gene derived from a

XX CC mutant of coryneform bacterium, measuring expression amount and

XX CC analysing the expression profile or expression pattern of a gene derived

XX CC from Coryneform bacterium, and identifying a homologue of a gene derived

XX CC from coryneform bacterium. Coryneform bacteria are useful for producing

XX CC particularly L-lysine. The present sequence is a protein described

XX CC in the exemplification of the invention.

XX CC Note: The sequence data for this patent did not form part of the printed

XX CC specification, but was obtained in electronic format directly from the

XX CC European Patent Office.

XX SQ Sequence 592 AA;

Query Match

Best Local Similarity 34.1%; Score 44; DB 22; Length 592;

Matches 10; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

OY 1 SDALEALADQTDALQSEEAAYV 22

DB 442 SDAMNALTTRDASSERLPVI 463

|||: ||: ||: ||:

## RESULT 34

AAW81028

ID AAW81028 standard; Protein: 951 AA.

XX AC AAW81028;

XX DT 12-MAY-1999 (first entry)

XX DE ER interacting domain of AIB1 protein.

XX KW AIB1; amplified in breast cancer; cancer; steroid; receptor;

XX KW coactivator; SCR; estrogen; ER; estrogen dependent transcription;

XX KW breast cancer; lung cancer; colon cancer; prostate cancer;

XX KW melanoma; PAS region.

XX OS Homo sapiens.

XX PN W09857982-A2.

XX PD 23-DEC-1998.

XX PF 17-JUN-1998; 98WO-US12689.

XX PR 17-JUN-1997; 97US-0049728.

XX PA (USGO ) US GOVERNMENT.

XX PI Meltzer P, Trent JM;

XX DR WPI; 1999-080946/07.

XX PT New isolated steroid receptor co-activator, AIB1 - used to develop

XX PT products for the diagnosis and treatment of steroid-responsive

XX PT tumours, e.g. breast, lung, prostate or colon cancers or melanomas

XX PS Claim 6; Page 34-35; 57pp; English.

XX CC The AIB1 protein is a member of the steroid receptor coactivator-1

XX CC (SCR-1) family of nuclear receptor co-activators that interact with

XX CC estrogen receptors (ER) to enhance ER-dependent transcription.

XX CC The AIB1 gene is amplified and over-expressed in certain cancers.

XX CC In particular breast cancer and steroid hormone responsive cancers.

XX CC The AIB1 polypeptide can be used to identify compounds which inhibit

XX CC ER-dependent transcription. Increased expression of the AIB1 gene

XX CC indicates aberrantly proliferating cells, thus detection of

XX CC increased expression of the AIB1 gene or an increase in the number

XX CC of copies of the AIB1 gene can be used to diagnose cancer or a

XX CC predisposition towards developing cancer. Compounds which inhibit

XX CC expression of AIB1 or compounds which inhibit interaction of AIB1

XX CC with steroid receptors or nuclear co-factors can be used for

XX CC reducing the proliferation of cancer cells. The PAS domain of the

XX CC AIB protein is a highly conserved domain among the SRC-1 family of

XX CC proteins. The ER-interacting domain of AIB1 mediates binding of

XX CC AIB1 with a steroid hormone receptor protein.

XX SQ Sequence 951 AA;

Query Match

Best Local Similarity 34.1%; Score 44; DB 20; Length 951;

Matches 10; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

OY 5 EALADQTDALQSEEAAYVVKADNA 27

DB 787 QALEPKQDAFGQEAAYVMDQKA 809

:||: ||: ||: |||:

## RESULT 35

AAB59278

ID AAB59278 standard; protein: 1412 AA.

XX AC AAB59278;

XX DT 27-MAR-2001 (first entry)





```

XX Example 5; Page 59; 90pp; English.
XX The present sequence is that of a homomeric coiled-coil domain
XX derived from amino acids 249-281 of Saccharomyces cerevisiae GCN4
XX (general control of nitrogen). The native GCN4 sequence was
XX modified to include a caspase site. A polypeptide useful in the
XX present invention is capable of multimerising with similar or
XX different polypeptides. The polypeptide requires a binding domain,
XX such as a coiled-coil domain, which allows multimerisation. The
XX multimer preferably comprises a first polypeptide associated with a
XX label and a second polypeptide, where (a) at least 1 of the
XX polypeptides is susceptible to protease digestion, (b) association
XX of the polypeptides to form a multimer is detectable via a signal
XX emitted by the signal, and (c) digestion of at least 1 polypeptides
XX results in dissociation of the multimer and modulation of the signal
XX emitted by the label. The multimer is used to monitor the activity
XX of a protease. In an example of the method, the polypeptide
XX incorporates the present GCN4 coiled-coil domain and is used in a
XX caspase assay, with the signal analysed by fluorescence resonance
XX energy transfer.
XX SQ Sequence 33 AA;
    Query Match 33.3%; Score 43; DB 21; Length 33;
    Best Local Similarity 40.9%; Pred. No. 18;
    Matches 9; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

QY 2 DALEALADQTDALQSEEAAYVK 23
   | : | | | : | : | : |
Db 7 DKVEELLNDHNEVAACLKK 28

RESULT 40
AAY44712
ID AAY44712 standard; Protein; 92 AA.
XX
XX AC AAY44712;
XX
XX DT 25-APR-2000 (first entry)
XX
XX DE Shorthorn sculpin skin-type antifreeze protein.
XX
XX KW clone s3-2; shorthorn sculpin skin-type antifreeze protein; sssAFP;
XX antibacterial; cold tolerance; ice recrystallisation;
XX refrigerated food; antifreeze.
XX
XX OS Myoxocephalus scorpius.
XX
XX PN WO200000512-A2.
XX
XX PD 06-JAN-2000.
XX
XX PF 25-JUN-1999; 99WO-CA00601.
XX
XX PR 26-JUN-1998; 98US-0090794.
XX 07-AUG-1998; 98US-0095713.
XX 24-JUN-1999; 99US-0344529.
XX
XX PA (HSCR-) HSC RES & DEV LP.
XX
XX PI Hew CL;
XX
XX DR WPI: 2000-170905/15.
XX DR N-PSDB; AA249960.
XX
XX PT Novel antifreeze polypeptides and polynucleotides used to make cells
XX cold resistant and to improve the palatability of cold foods and
XX liquids.
XX
XX PS Example 1; Fig 1; 61pp; English.
XX
XX CC The present sequence is shorthorn sculpin skin-type antifreeze protein

```

```

CC (sssAFP-2) encoded by cDNA clone s3-2. The AFPs are alanine-rich
CC polypeptides that are synthesised in the peripheral tissues, such as
CC skin and gills, and serve to depress the freezing temperature of
CC intracellular fluids. The AFPs are used to make aqueous compositions
CC resistant to freezing, to improve the cold tolerance, for inhibiting ice
CC recrystallisation, and for stabilising biological membranes. They are
CC also used to improve the shelf life and food quality of many
CC refrigerated foods and have antibacterial properties. This has many
CC useful applications in medicine, agriculture and aquaculture.
XX
XX SQ Sequence 92 AA;
    Query Match 33.3%; Score 43; DB 21; Length 92;
    Best Local Similarity 39.3%; Pred. No. 59;
    Matches 11; Conservative 5; Mismatches 12; Indels 0; Gaps 0;

QY 1 SDALEALADQTDALQSEEAAYVKADNAA 28
   | | | | : | : | | | | |
Db 32 SAAAAAIAIAEAAEAEEAAATKSNVA 59

```

Search completed: October 13, 2002, 02:08:16  
 Job time : 13.3797 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2002, 02:06:51 ; Search time 24.6578 Seconds  
(without alignments)  
157.503 Million cell updates/sec

Title: US-09-847-539A-6

Perfect score: 777

Sequence: 1 VDSIEQPIRIIPNGGTLNL.....KKEERQNVNLTPTTGEENP 159

Scoring table: BLOSUM62

Gapop 60.0 , Gapext 60.0

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents\_AA:\*

- 1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*
- 2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*
- 3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep.\*
- 4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep.\*
- 5: /cgn2\_6/ptodata/2/iaa/PT05\_COMB.pep.\*
- 6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	91	11.7	413	US-08-669-408B-10	Sequence 10, Appl
2	85	10.9	664	US-08-669-408B-2	Sequence 2, Appl
3	76	9.8	50	US-08-294-189-14	Sequence 14, Appl
4	76	9.8	3248	US-08-353-700-1	Sequence 1, Appl
5	76	9.8	3248	PCT-US95-16216-1	Sequence 1, Appl
6	75	9.7	2482	US-08-328-254-6	Sequence 6, Appl
7	69	8.9	46	US-08-363-311-25	Sequence 25, Appl
8	69	8.9	46	US-08-463-288A-25	Sequence 25, Appl
9	69	8.9	46	US-08-470-445A-25	Sequence 25, Appl
10	69	8.9	46	US-08-462-679-25	Sequence 25, Appl
11	69	8.9	46	US-08-466-210A-25	Sequence 25, Appl
12	69	8.9	46	US-08-467-147A-25	Sequence 25, Appl
13	69	8.9	46	US-08-469-014-25	Sequence 25, Appl
14	69	8.9	46	US-09-346-290-25	Sequence 25, Appl
15	69	8.9	46	PCT-US93-10506A-25	Sequence 25, Appl
16	69	8.9	46	PCT-US93-10506-25	Sequence 25, Appl
17	68	8.8	1786	US-08-973-463-8	Sequence 8, Appl
18	67	8.6	472	US-08-216-894-10	Sequence 10, Appl
19	67	8.6	472	US-09-115-746-10	Sequence 10, Appl
20	66	8.5	46	US-08-363-311-26	Sequence 26, Appl
21	66	8.5	46	US-08-463-288A-26	Sequence 26, Appl
22	66	8.5	46	US-08-470-445A-26	Sequence 26, Appl
23	66	8.5	46	US-08-462-679-26	Sequence 26, Appl
24	66	8.5	46	US-08-466-210A-26	Sequence 26, Appl
25	66	8.5	46	US-08-467-147A-26	Sequence 26, Appl
26	66	8.5	46	US-08-469-014-26	Sequence 26, Appl
27	66	8.5	46	US-09-346-290-26	Sequence 26, Appl

28	8.5	46	5	PCT-US93-10506A-26	Sequence 26, Appl
29	66	8.5	46	PCT-US93-10506-26	Sequence 26, Appl
30	66	8.5	593	US-08-591-079-8	Sequence 8, Appl
31	66	8.5	593	US-08-591-079-10	Sequence 10, Appl
32	65	8.4	1162	US-08-728-323A-2	Sequence 2, Appl
33	64	8.2	564	US-08-216-894-2	Sequence 2, Appl
34	64	8.2	564	US-09-115-746-2	Sequence 2, Appl
35	64	8.2	643	US-08-216-894-8	Sequence 8, Appl
36	64	8.2	643	US-09-115-746-8	Sequence 8, Appl
37	64	8.2	1111	US-08-317-450B-15	Sequence 15, Appl
38	64	8.2	1111	US-08-800-593-15	Sequence 15, Appl
39	64	8.2	1193	US-08-317-450B-13	Sequence 13, Appl
40	64	8.2	1193	US-08-800-593-13	Sequence 13, Appl
41	63	8.1	288	US-08-961-858-6	Sequence 6, Appl
42	63	8.1	288	US-09-089-593-6	Sequence 6, Appl
43	63	8.1	288	US-08-950-925-4	Sequence 4, Appl
44	62	8.0	461	US-09-346-408-8	Sequence 8, Appl
45	62	8.0	999	US-08-770-301A-3	Sequence 3, Appl

ALIGNMENTS

RESULT 1  
US-08-669-408B-10  
; Sequence 10, Application US/08669408B  
; Patent No. 6100055  
; GENERAL INFORMATION:  
; APPLICANT: GUSS, Bengt  
; APPLICANT: JONSSON, Hans  
; APPLICANT: LINDBERG, Martin  
; APPLICANT: MUELLER, Hans-Peter  
; APPLICANT: RANTAMAKI, Liisa K.  
; TITLE OF INVENTION: METHOD AND MEANS FOR PRODUCING  
; TITLE OF INVENTION: PLASMAPROTEINASE INHIBITOR-BINDING PROTEINS  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSER: Foley & Lardner  
; STREET: 3000 K Street, N.W., Suite 500  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20007-5109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC Compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/669,408B  
; FILING DATE: 03-JUL-1996  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: WO PCT/SE94/00826  
; FILING DATE: 06-SEP-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: SE 9302855-3  
; FILING DATE: 06-SEP-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BENT, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 61743/102  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202)672-5300  
; TELEFAX: (202)672-5399  
; TELEX: 904136  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 413 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-669-408B-10





COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/353,700  
FILING DATE: 09-DEC-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: REED, JANET E.  
REGISTRATION NUMBER: 36,252  
TELEPHONE: (215) 563-4100  
TELEFAX: (215) 563-4044  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3248 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: HUMAN  
US-08-353-700-1

Query Match 9.8%; Score 76; DB 1; Length 3248;  
Best Local Similarity 21.0%; Pred. No. 2.2;  
Matches 21; Conservative 28; Mismatches 51; Indels 0; Gaps 0;  
QY 46 EDKEATTATEAASSDALEALADOTDALQSEAAVVKADNNAASDALEALADOTDALQSEEA 105  
DB 2520 EKEQEKVQMKESSTAMEMLOTQKELNERNVAALHNDQKACKKEQNLSSQVECLELEKA 2579  
QY 106 EVVQSDNAASDAWEKAATPIALDVKKTKDKPKVVKKEERQ 145  
DB 2580 QLLQGLDEAKNNIVLQSSVKGLIQEVEDGKQKLEKDEE 2619

## RESULT 5

PCT-US95-16216-1  
Sequence 1, Application PC/TUS9516216  
GENERAL INFORMATION:  
APPLICANT: Yen, Timothy J.  
APPLICANT: Ratiner, Jerome B.  
TITLE OF INVENTION: Nucleic Acid Encoding a Transiently  
Expressed Kinetochores Protein, and Methods of Use  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dann, Dorfman, Herrell and Skillman  
STREET: 1601 Market Street Suite 720  
CITY: Philadelphia  
STATE: PA  
COUNTRY: USA  
ZIP: 19103-2307  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/16216  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/353,700  
FILING DATE: 09-DEC-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Reed, Janet E.  
REGISTRATION NUMBER: 36,252  
TELEPHONE: (215) 563-4100  
TELEFAX: (215) 563-4044

INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3248 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: not relevant  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
PCT-US95-16216-1

Query Match 9.8%; Score 76; DB 5; Length 3248;  
Best Local Similarity 21.0%; Pred. No. 2.2;  
Matches 21; Conservative 28; Mismatches 51; Indels 0; Gaps 0;  
QY 46 EDKEATTATEAASSDALEALADOTDALQSEAAVVKADNNAASDALEALADOTDALQSEEA 105  
DB 2520 EKEQEKVQMKESSTAMEMLOTQKELNERNVAALHNDQKACKKEQNLSSQVECLELEKA 2579  
QY 106 EVVQSDNAASDAWEKAATPIALDVKKTKDKPKVVKKEERQ 145  
DB 2580 QLLQGLDEAKNNIVLQSSVKGLIQEVEDGKQKLEKDEE 2619

## RESULT 6

US-08-328-254-6  
Sequence 6, Application US/08328254  
Patent No. 5710022  
GENERAL INFORMATION:  
APPLICANT: Zhu, Xueliang  
APPLICANT: Lee, Wen-Hwa  
TITLE OF INVENTION: A No. 5710022el Nuclear Mitotic Phosphoprotein  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Campbell and Flores  
STREET: 4370 La Jolla Village Drive, Suite 700  
CITY: San Diego  
STATE: California  
COUNTRY: USA  
ZIP: 92122  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/328,254  
FILING DATE: 24-OCT-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/141,239  
FILING DATE: 22-OCT-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Campbell, Cathryn A.  
REGISTRATION NUMBER: 31,815  
REFERENCE/DOCKET NUMBER: P-CJ 1191  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 535-9001  
TELEFAX: (619) 535-8949  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2482 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-328-254-6

Query Match 9.7%; Score 75; DB 1; Length 2482;  
Best Local Similarity 21.0%; Pred. No. 2.1;  
Matches 21; Conservative 28; Mismatches 51; Indels 0; Gaps 0;  
QY 46 EDKEATTATEAASSDALEALADOTDALQSEAAVVKADNNAASDALEALADOTDALQSEEA 105



```
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/470,445A
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/363,311
; FILING DATE: 22-DEC-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/968,866
; FILING DATE: 02-NOV-1992
; APPLICATION DATA:
; APPLICATION NUMBER: US 07/408,036
; FILING DATE: 15-SEP-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Bugalsky, Lawrence B.
; REGISTRATION NUMBER: 35,086
; REFERENCE/DOCKET NUMBER: 0609.237000A
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; TELEX: 248636 SSK
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 46 amino acids
; TYPE: amino acid
; TOPOLOGY: both
; MOLECULE TYPE: peptide
; US-08-470-445A-25

Query Match      8.9%; Score 69; DB 2; Length 46;
Best Local Similarity 65.0%; Pred. No. 0.075;
Matches 13; Conservative 2; Mismatches 5; Indels 5; Gaps 0;

OY 140 KKEERQNVNLTPTTGESNP 159
   |||: : ||||| |||
Db 2 KEDAKKAETLPTTGESNP 21

RESULT 10
US-08-462-679-25
; Sequence 25, Application US/08462679
; Patent No. 5847081
; GENERAL INFORMATION:
; APPLICANT: Michel, James L.
; APPLICANT: Kasper, Dennis L.
; APPLICANT: Ausubel, Frederick M.
; APPLICANT: Madoff, Lawrence C.
; TITLE OF INVENTION: Conjugate Vaccine For Group B
; NUMBER OF SEQUENCES: 65
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, NW, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/462,679
; FILING DATE: 05-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/363,311
; FILING DATE: 22-DEC-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/968,866
; FILING DATE: 02-NOV-1992
```

```
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/408,036
; FILING DATE: 15-SEP-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Bugalsky, Lawrence B.
; REGISTRATION NUMBER: 35,086
; REFERENCE/DOCKET NUMBER: 0609.2370008
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; TELEX: 248636 SSK
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 46 amino acids
; TYPE: amino acid
; TOPOLOGY: both
; MOLECULE TYPE: peptide
; US-08-462-679-25

Query Match      8.9%; Score 69; DB 2; Length 46;
Best Local Similarity 65.0%; Pred. No. 0.075;
Matches 13; Conservative 2; Mismatches 5; Indels 5; Gaps 0;

OY 140 KKEERQNVNLTPTTGESNP 159
   |||: : ||||| |||
Db 2 KEDAKKAETLPTTGESNP 21

RESULT 11
US-08-466-210A-25
; Sequence 25, Application US/08466210A
; Patent No. 5858362
; GENERAL INFORMATION:
; APPLICANT: Michel, James L.
; APPLICANT: Kasper, Dennis L.
; APPLICANT: Ausubel, Frederick M.
; APPLICANT: Madoff, Lawrence C.
; TITLE OF INVENTION: Conjugate Vaccine For Group B
; NUMBER OF SEQUENCES: 65
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, NW, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/466,210A
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/363,311
; FILING DATE: 22-DEC-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/968,866
; FILING DATE: 02-NOV-1992
; APPLICATION DATA:
; APPLICATION NUMBER: US 07/408,036
; FILING DATE: 15-SEP-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Bugalsky, Lawrence B.
; REGISTRATION NUMBER: 35,086
; REFERENCE/DOCKET NUMBER: 0609.237000B
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; TELEX: 248636 SSK
```

; INFORMATION FOR SEQ ID NO: 25:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 46 amino acids  
; TYPE: amino acid  
; TOPOLOGY: both  
; MOLECULE TYPE: peptide  
US-08-466-210A-25

Query Match 8.9%; Score 69; DB 2; Length 46;  
Best Local Similarity 65.0%; Pred. No. 0.075;  
Matches 13; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 140 KKEERQNVNTLPTTGESNP 159  
Db 2 KKEDAKKAETLPTTGESNP 21  
||||: ||||||| |||

## RESULT 12

US-08-467-147A-25  
; Sequence 25, Application US/08467147A  
; Patent No. 5908629

; GENERAL INFORMATION:  
; APPLICANT: Michel, James L.  
; APPLICANT: Kasper, Dennis L.  
; APPLICANT: Ausubel, Frederick C.  
; APPLICANT: Madoff, Lawrence C.  
; TITLE OF INVENTION: Conjugate Vaccine For Group B  
; TITLE OF INVENTION: Conjugate Vaccine For Group B  
; NUMBER OF SEQUENCES: 65  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.  
; STREET: 1100 New York Avenue, NW, Suite 600  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20005-3934

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/467,147A  
; FILING DATE: 06-JUN-1995  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/363,311  
; FILING DATE: 22-DEC-1994

; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/968,866  
; FILING DATE: 02-NOV-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/408,036  
; FILING DATE: 15-SEP-1989

; ATTORNEY/AGENT INFORMATION:  
; NAME: Bugalsky, Lawrence B.  
; REGISTRATION NUMBER: 35,086  
; REFERENCE/DOCKET NUMBER: 0609.2370009  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 371-2600  
; TELEFAX: (202) 371-2540  
; TELEX: 248636 SSK

; INFORMATION FOR SEQ ID NO: 25:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 46 amino acids  
; TYPE: amino acid  
; TOPOLOGY: both  
; MOLECULE TYPE: peptide  
US-08-467-147A-25

Query Match 8.9%; Score 69; DB 2; Length 46;  
Best Local Similarity 65.0%; Pred. No. 0.075;  
Matches 13; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 140 KKEERQNVNTLPTTGESNP 159  
Db 2 KKEDAKKAETLPTTGESNP 21  
||||: ||||||| |||

## RESULT 13

US-08-469-014-25  
; Sequence 25, Application US/08469014  
; Patent No. 5968521

; GENERAL INFORMATION:  
; APPLICANT: Michel, James L.  
; APPLICANT: Kasper, Dennis L.  
; APPLICANT: Ausubel, Frederick C.  
; APPLICANT: Madoff, Lawrence C.  
; TITLE OF INVENTION: Conjugate Vaccine Against Group B  
; TITLE OF INVENTION: Conjugate Vaccine Against Group B  
; NUMBER OF SEQUENCES: 65  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.  
; STREET: 1100 New York Avenue, NW, Suite 600  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20005-3934  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/469,014  
; FILING DATE: 05-JUN-1995  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/363,311  
; FILING DATE: 22-DEC-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/968,866  
; FILING DATE: 02-NOV-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/408,036  
; FILING DATE: 15-SEP-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Bugalsky, Lawrence B.  
; REGISTRATION NUMBER: 35,086  
; REFERENCE/DOCKET NUMBER: 0609.2370006  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 371-2600  
; TELEFAX: (202) 371-2540  
; TELEX: 248636 SSK  
; INFORMATION FOR SEQ ID NO: 25:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 46 amino acids  
; TYPE: amino acid  
; TOPOLOGY: both  
; MOLECULE TYPE: peptide  
US-08-469-014-25

Query Match 8.9%; Score 69; DB 2; Length 46;  
Best Local Similarity 65.0%; Pred. No. 0.075;  
Matches 13; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 140 KKEERQNVNTLPTTGESNP 159  
Db 2 KKEDAKKAETLPTTGESNP 21  
||||: ||||||| |||

## RESULT 14

US-09-346-290-25

; Sequence 25, Application US/09346290  
; Patent No. 6342223  
; GENERAL INFORMATION:

Query Match 8.9%; Score 69; DB 2; Length 46;  
Best Local Similarity 65.0%; Pred. No. 0.075;  
Matches 13; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

```
; APPLICANT: Michel, James L.
; APPLICANT: Kasper, Dennis L.
; APPLICANT: Ausubel, Frederick M.
; APPLICANT: Madoff, Lawrence C.
; TITLE OF INVENTION: Conjugate vaccine Against Group B
; NUMBER OF SEQUENCES: 65
; CORRESPONDENCE ADDRESS:
; STREET: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/346,290
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/469,014
; FILING DATE:
; APPLICATION NUMBER: US 07/968,866
; FILING DATE: 02-NOV-1992
; APPLICATION NUMBER: US 07/408,036
; FILING DATE: 15-SEP-1989
; APPLICATION NUMBER:
; ATTORNEY/AGENT INFORMATION:
; NAME: Bugalsky, Lawrence B.
; REGISTRATION NUMBER: 35,086
; REFERENCE/DOCKET NUMBER: 0609.2370006
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2540
; TELEFAX: (202) 371-2540
; TELEX: 248636 SSK
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 46 amino acids
; TYPE: amino acid
; TOPOLOGY: both
; MOLECULE TYPE: peptide
; US-09-346-290-25

Query Match      8.9%; Score 69; DB 4; Length 46;
Best Local Similarity 65.0%; Pred. No. 0.075;
Matches 13; Conservative 2; Mismatches 5; Indels 5; Gaps 0;

QY 140 KKEERQNVNLTPTTGESNP 159
   ||| : ||||| |||
Db 2 KKEDAKKAETLPTTGESNP 21

RESULT 15
PCT-US93-10506A-25
; Sequence 25, Application PC/TUS9310506A
; GENERAL INFORMATION:
; APPLICANT: THE GENERAL HOSPITAL CORPORATION
; APPLICANT: Fruit Street
; APPLICANT: Boston, Massachusetts 02114
; APPLICANT: United States of America
; APPLICANT: 75 Francis Street
; APPLICANT: Boston, Massachusetts 02115
; APPLICANT: United States of America
; TITLE OF INVENTION: Conjugate Vaccine Against Group B
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
```

```
; STREET: 1100 New York Ave., Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/10506A
; FILING DATE: 02-NOV-1993
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/968,866
; FILING DATE: 02-NOV-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Cimbala, Michele A.
; REGISTRATION NUMBER: 33,851
; REFERENCE/DOCKET NUMBER: 0609.237PC01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; TELEX: 248636 SSK
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 46 amino acids
; TYPE: amino acid
; TOPOLOGY: both
; MOLECULE TYPE: peptide
; PCT-US93-10506A-25

Query Match      8.9%; Score 69; DB 5; Length 46;
Best Local Similarity 65.0%; Pred. No. 0.075;
Matches 13; Conservative 2; Mismatches 5; Indels 5; Gaps 0;

QY 140 KKEERQNVNLTPTTGESNP 159
   ||| : ||||| |||
Db 2 KKEDAKKAETLPTTGESNP 21

RESULT 16
PCT-US93-10506-25
; Sequence 25, Application PC/TUS9310506
; GENERAL INFORMATION:
; APPLICANT: Michel, James L.
; APPLICANT: Kasper, Dennis
; APPLICANT: Ausubel, Frederick M.
; APPLICANT: Madoff, Lawrence C.
; TITLE OF INVENTION: Conjugate Vaccine Against Group B
; TITLE OF INVENTION: Streptococcus
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1100 New York Avenue, N.W.; Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/10506
; FILING DATE: 02-NOV-1993
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/968,866
; FILING DATE: 02-NOV-1992
; ATTORNEY/AGENT INFORMATION:
```

```

Query Match      8.6%; Score 67; DB 2; Length 472;
Best Local Similarity 20.08; Pred.No. 2.6;
Matches 18; Conservative 20; Mismatches 52; Indels 0; Gaps
QY 33 EERAIDELKKQAIEDKATTAAIEAASSDALEALADOTDALQSEAAVVKADNAASDALEA 92
   |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Db 255 EKQRAAEATKVAAEAKQKAAEAATKVAAEKORAAEATKVAAEKQKAAEATKVAGDEKQK 314
QY 93 LADOTDALQSEAEVVQSDDNNAASDAWEKAA 122
   |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Db 315 AAETATKVAAEKQKAAEAATKVAAEAKQKAA 344

RESULT 19
US-09-115-746-10
; Sequence 10, Application US/09115746
; Patent No. 6228601
; GENERAL INFORMATION:
; APPLICANT: Kirchoff, Louis V.
; APPLICANT: Otsu, Keiko
; TITLE OF INVENTION: POLYPEPTIDES FOR DIAGNOSING INFECTION
; TITLE OF INVENTION: WITH TRYPA NOSOMA CRUZI
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/115,746
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/216,894
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 85326/102/DRLO

```

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 472 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-115-746-10

Query Match 8.5%; Score 67; DB 4; Length 472;  
Best Local Similarity 20.0%; Pred. No. 2.6;  
Matches 18; Conservative 20; Mismatches 52; Indels 0; Gaps 0;  
QY 33 EERAIDELKKQATEKEATTATTAASSDALEALADOTDALOSEAAVVKADNAASDALEA 92  
DB 255 EKORAAEATKVAEAKOKAAEATKVAEAKOKAAEATKVAEAKOKAAEATKVAAGDEKQK 314  
QY 93 LADOTDALOSEAAEVVQSDNAASDAEAKA 122  
DB 315 AAEATKVAEAKOKAAEATKVAEAKOKAA 344

RESULT 20  
US-08-363-311-26  
Sequence 26, Application US/08363311  
Patent No. 6548241  
GENERAL INFORMATION:  
APPLICANT: Michel, James L.  
APPLICANT: Ausubel, Frederick M.  
TITLE OF INVENTION: Conjugate Vaccine Against Group B  
NUMBER OF SEQUENCES: 29  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sterne, Kessler, Goldstein & Fox  
STREET: 1225 Connecticut Avenue, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20036-2678  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION NUMBER: US/08/363,311  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/968,866  
FILING DATE: 02-NOV-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Cimbala, Michele A.  
REGISTRATION NUMBER: 33,851  
REFERENCE/DOCKET NUMBER: 0609.3740004  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 466-0800  
TELEFAX: (202) 833-8716  
TELEX: 248636 SSK  
INFORMATION FOR SEQ ID NO: 26:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 46 amino acids  
TYPE: amino acid  
TOPOLOGY: both  
MOLECULE TYPE: peptide  
US-08-363-311-26

Query Match 8.5%; Score 66; DB 1; Length 46;  
Best Local Similarity 60.0%; Pred. No. 0.18;  
Matches 12; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 140 KKEERQNVNTLPTTGESNP 159  
DB 2 KDDAKKAETLPTTGESNP 21  
RESULT 21  
US-08-463-288A-26  
Sequence 26, Application US/08463288A  
Patent No. 5820860  
GENERAL INFORMATION:  
APPLICANT: Michel, James L.  
APPLICANT: Kasper, Dennis L.  
APPLICANT: Ausubel, Frederick M.  
APPLICANT: Madoff, Lawrence C.  
TITLE OF INVENTION: Conjugate Vaccine For Group B  
NUMBER OF SEQUENCES: 65  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.  
STREET: 1100 New York Avenue, NW, Suite 600  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005-3934  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION NUMBER: US/08/463,288A  
FILING DATE: 05-JUN-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/363,311  
FILING DATE: 22-DEC-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/968,866  
FILING DATE: 02-NOV-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/408,036  
FILING DATE: 15-SEP-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: Bugaisky, Lawrence B.  
REGISTRATION NUMBER: 35,086  
REFERENCE/DOCKET NUMBER: 0609.2370007  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 371-2600  
TELEFAX: (202) 371-2540  
TELEX: 248636 SSK  
INFORMATION FOR SEQ ID NO: 26:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 46 amino acids  
TYPE: amino acid  
TOPOLOGY: both  
MOLECULE TYPE: peptide  
US-08-463-288A-26

Query Match 8.5%; Score 66; DB 2; Length 46;  
Best Local Similarity 60.0%; Pred. No. 0.18;  
Matches 12; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 140 KKEERQNVNTLPTTGESNP 159  
DB 2 KDDAKKAETLPTTGESNP 21

RESULT 22  
US-08-470-445A-26  
Sequence 26, Application US/08470445A  
Patent No. 5843444  
GENERAL INFORMATION:

APPLICANT: Michel, James L.  
APPLICANT: Kasper, Dennis L.  
APPLICANT: Ausubel, Frederick M.  
APPLICANT: Madoff, Lawrence C.  
TITLE OF INVENTION: Conjugate Vaccine Against Group B  
TITLE OF INVENTION: Streptococcus  
NUMBER OF SEQUENCES: 65  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.  
STREET: 1100 New York Avenue, NW, Suite 600  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005-3934  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/470.445A  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/363.311  
FILING DATE: 22-DEC-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/968.866  
FILING DATE: 02-NOV-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Bugalsky, Lawrence B.  
REGISTRATION NUMBER: 35,086  
REFERENCE/DOCKET NUMBER: 0609.237000A  
TELEPHONE: (202) 371-2540  
TELEFAX: (202) 371-2540  
TELEX: 248636 SSK  
INFORMATION FOR SEQ ID NO: 26:  
APPLICATION NUMBER: US 07/408.036  
FILING DATE: 15-SEP-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: Bugalsky, Lawrence B.  
REGISTRATION NUMBER: 35,086  
REFERENCE/DOCKET NUMBER: 0609.237000A  
TELEPHONE: (202) 371-2540  
TELEFAX: (202) 371-2540  
TELEX: 248636 SSK  
INFORMATION FOR SEQ ID NO: 26:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 46 amino acids  
TYPE: amino acid  
TOPOLOGY: both  
MOLECULE TYPE: peptide  
US-08-470-445A-26

Query Match 8.5%; Score 66; DB 2; Length 46;  
Best Local Similarity 60.0%; Pred. No. 0.18;  
Matches 12; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 140 KKEQRQNTLPTTGEESNP 159  
Db 2 KKDDAKKAETLPTTGEESNP 21  
|||: : ||||| |||

RESULT 23  
US-08-462-679-26  
Sequence 26, Application US/08462679  
Patent No. 5847081  
GENERAL INFORMATION:  
APPLICANT: Michel, James L.  
APPLICANT: Kasper, Dennis L.  
APPLICANT: Ausubel, Frederick M.  
APPLICANT: Madoff, Lawrence C.  
TITLE OF INVENTION: Conjugate Vaccine For Group B  
TITLE OF INVENTION: Streptococcus  
NUMBER OF SEQUENCES: 65  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.  
STREET: 1100 New York Avenue, NW, Suite 600  
CITY: Washington

STATE: D.C.  
COUNTRY: USA  
ZIP: 20005-3934  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/462.679  
FILING DATE: 05-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/363.311  
FILING DATE: 22-DEC-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/968.866  
FILING DATE: 02-NOV-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Bugalsky, Lawrence B.  
REGISTRATION NUMBER: 35,086  
REFERENCE/DOCKET NUMBER: 0609.2370008  
TELEPHONE: (202) 371-2600  
TELEFAX: (202) 371-2540  
TELEX: 248636 SSK  
INFORMATION FOR SEQ ID NO: 26:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 46 amino acids  
TYPE: amino acid  
TOPOLOGY: both  
MOLECULE TYPE: peptide  
US-08-462-679-26

Query Match 8.5%; Score 66; DB 2; Length 46;  
Best Local Similarity 60.0%; Pred. No. 0.18;  
Matches 12; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 140 KKEQRQNTLPTTGEESNP 159  
Db 2 KKDDAKKAETLPTTGEESNP 21  
|||: : ||||| |||

RESULT 24  
US-08-466-210A-26  
Sequence 26, Application US/08466210A  
Patent No. 5858362  
GENERAL INFORMATION:  
APPLICANT: Michel, James L.  
APPLICANT: Kasper, Dennis L.  
APPLICANT: Ausubel, Frederick M.  
APPLICANT: Madoff, Lawrence C.  
TITLE OF INVENTION: Conjugate Vaccine For Group B  
TITLE OF INVENTION: Streptococcus  
NUMBER OF SEQUENCES: 65  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.  
STREET: 1100 New York Avenue, NW, Suite 600  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005-3934  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/466.210A  
FILING DATE: 06-JUN-1995



CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
FILING DATE: 22-DEC-1994  
PRIOR APPLICATION NUMBER: US 08/363,311  
FILING DATE: 02-NOV-1992  
PRIOR APPLICATION NUMBER: US 07/968,866  
FILING DATE: 02-NOV-1992  
PRIOR APPLICATION NUMBER: US 07/408,036  
FILING DATE: 15-SEP-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: Bugalsky, Lawrence B.  
REGISTRATION NUMBER: 35,086  
REFERENCE/DOCKET NUMBER: 0609,237000B  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 371-2600  
TELEFAX: (202) 371-2540  
INFORMATION FOR SEQ ID NO: 26:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 46 amino acids  
TYPE: amino acid  
TOPOLOGY: both  
MOLECULE TYPE: peptide  
US-08-466-210A-26

Query Match 8.5%; Score 66; DB 2; Length 46;  
Best Local Similarity 60.0%; Pred. No. 0.16;  
Matches 12; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 140 KKEERQNVNTLPTTGESNP 159  
|||: : ||||| |||  
Db 2 KKDDAKKAEUPTTGGESNP 21

RESULT 25  
US-08-467-147A-26  
Sequence 26, Application US/08467147A  
Patent No. 5908629  
GENERAL INFORMATION:  
APPLICANT: Michel, James L.  
APPLICANT: Kasper, Dennis L.  
APPLICANT: Ausubel, Frederick M.  
APPLICANT: Madoff, Lawrence C.  
TITLE OF INVENTION: Conjugate Vaccine For Group B  
NUMBER OF SEQUENCES: 65  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.  
STREET: 1100 New York Avenue, NW, Suite 600  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005-3934  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/467,147A  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/363,311  
FILING DATE: 22-DEC-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/968,866  
FILING DATE: 02-NOV-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/408,036  
FILING DATE: 15-SEP-1989  
ATTORNEY/AGENT INFORMATION:

NAME: Bugalsky, Lawrence B.  
REGISTRATION NUMBER: 35,086  
REFERENCE/DOCKET NUMBER: 0609,2370009  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 371-2600  
TELEFAX: (202) 371-2540  
INFORMATION FOR SEQ ID NO: 26:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 46 amino acids  
TYPE: amino acid  
TOPOLOGY: both  
MOLECULE TYPE: peptide  
US-08-467-147A-26

Query Match 8.5%; Score 66; DB 2; Length 46;  
Best Local Similarity 60.0%; Pred. No. 0.16;  
Matches 12; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 140 KKEERQNVNTLPTTGESNP 159  
|||: : ||||| |||  
Db 2 KKDDAKKAEUPTTGGESNP 21

RESULT 26  
US-08-469-014-26  
Sequence 26, Application US/08469014  
Patent No. 5968521  
GENERAL INFORMATION:  
APPLICANT: Michel, James L.  
APPLICANT: Kasper, Dennis L.  
APPLICANT: Ausubel, Frederick M.  
APPLICANT: Madoff, Lawrence C.  
TITLE OF INVENTION: Conjugate Vaccine Against Group B  
NUMBER OF SEQUENCES: 65  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.  
STREET: 1100 New York Avenue, NW, Suite 600  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005-3934  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/469,014  
FILING DATE: 05-JUN-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/363,311  
FILING DATE: 22-DEC-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/968,866  
FILING DATE: 02-NOV-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/408,036  
FILING DATE: 15-SEP-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: Bugalsky, Lawrence B.  
REGISTRATION NUMBER: 35,086  
REFERENCE/DOCKET NUMBER: 0609,2370006  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 371-2600  
TELEFAX: (202) 371-2540  
INFORMATION FOR SEQ ID NO: 26:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 46 amino acids  
TYPE: amino acid

; TOPOLOGY: both  
; MOLECULE TYPE: peptide  
US-08-469-014-26

Query Match 8.5%; Score 66; DB 2; Length 46;  
Best Local Similarity 60.0%; Pred. No. 0.18;  
Matches 12; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 140 KKEERQNVNLTPTTGESNP 159  
|::: | | | | | | | | | |  
Db 2 KKDDAKKAETLPTTGESNP 21

## RESULT 27

US-09-346-290-26  
; Sequence 26, Application US/09346290  
; Patent No. 6342223  
; GENERAL INFORMATION:  
; APPLICANT: Michel, James L.  
; APPLICANT: Kasper, Dennis L.  
; APPLICANT: Ausubel, Frederick M.  
; APPLICANT: Madoff, Lawrence C.  
; TITLE OF INVENTION: Conjugate Vaccine Against Group B  
; TITLE OF INVENTION: Streptococcus  
; NUMBER OF SEQUENCES: 65  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.  
; STREET: 1100 New York Avenue, NW, Suite 600  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20005-3934

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/346,290  
FILING DATE:

CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/469,014  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/968,866  
FILING DATE: 02-NOV-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/408,036  
FILING DATE: 13-SEP-1989

ATTORNEY/AGENT INFORMATION:  
NAME: Bugalsky, Lawrence B.  
REGISTRATION NUMBER: 35,086  
REFERENCE/DOCKET NUMBER: 0609.2370006  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 371-2600  
TELEFAX: (202) 371-2540  
INFORMATION FOR SEQ ID NO: 26:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 46 amino acids  
TYPE: amino acid

TOPOLOGY: both  
MOLECULE TYPE: peptide  
US-09-346-290-26

Query Match 8.5%; Score 66; DB 4; Length 46;  
Best Local Similarity 60.0%; Pred. No. 0.18;  
Matches 12; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 140 KKEERQNVNLTPTTGESNP 159  
|::: | | | | | | | | | |  
Db 2 KKDDAKKAETLPTTGESNP 21

## RESULT 28

PCT-US93-10506A-26  
; Sequence 26, Application PC/TUS9310506A  
; GENERAL INFORMATION:  
; APPLICANT: THE GENERAL HOSPITAL CORPORATION  
; APPLICANT: Fruit Street  
; APPLICANT: Boston, Massachusetts 02114  
; APPLICANT: United States of America  
; APPLICANT: 75 Francis Street  
; APPLICANT: Boston, Massachusetts 02115  
; APPLICANT: United States of America  
; TITLE OF INVENTION: Conjugate Vaccine Against Group B  
; TITLE OF INVENTION: Streptococcus  
; NUMBER OF SEQUENCES: 29  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox  
; STREET: 1100 New York Ave., Suite 600  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20005

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/10506A  
FILING DATE: 02-NOV-1993

CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/968,866  
FILING DATE: 02-NOV-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Cimbala, Michele A.  
REGISTRATION NUMBER: 33,851  
REFERENCE/DOCKET NUMBER: 0609.237PC01  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 371-2600  
TELEFAX: (202) 371-2540

INFORMATION FOR SEQ ID NO: 26:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 46 amino acids  
TYPE: amino acid  
TOPOLOGY: both  
MOLECULE TYPE: peptide  
PCT-US93-10506A-26

Query Match 8.5%; Score 66; DB 5; Length 46;  
Best Local Similarity 60.0%; Pred. No. 0.18;  
Matches 12; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 140 KKEERQNVNLTPTTGESNP 159  
|::: | | | | | | | | | |  
Db 2 KKDDAKKAETLPTTGESNP 21

## RESULT 29

PCT-US93-10506-26  
; Sequence 26, Application PC/TUS9310506  
; GENERAL INFORMATION:  
; APPLICANT: Michel, James L.  
; APPLICANT: Kasper, Dennis  
; APPLICANT: Ausubel, Frederick M.  
; APPLICANT: Madoff, Lawrence C.  
; TITLE OF INVENTION: Conjugate Vaccine Against Group B  
; TITLE OF INVENTION: Streptococcus  
; NUMBER OF SEQUENCES: 29  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/10506
FILING DATE: 02-NOV-1993
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/968,866
FILING DATE: 02-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: Cimbala, Michele A.
REGISTRATION NUMBER: 33,851
REFERENCE/DOCKET NUMBER: 0609.237PC01
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
TELEX: 248636 SSK
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 46 amino acids
TYPE: amino acid
TOPOLOGY: both
MOLECULE TYPE: peptide
PCT-US93-10506-26

```

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
84

STATE: DC  
COUNTRY: USA  
ZIP: 20006-1812  
COMPUTER READABLE FORM:

[illegible]

```
Db 199 KEALDKATDATVKGAGTDAKAEKAD 224
      || | | | | | | | | | |
RESULT 32
US-08-728-323A-2
; Sequence 2, Application US/08728323A
; Patent No. 5948676
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Bohenzky, Roy A.
; APPLICANT: Russo, James J.
; APPLICANT: Edelman, Isidore S.
; APPLICANT: Moore, Patrick S.
; TITLE OF INVENTION: Immediate Early Protein From Kaposi's
; TITLE OF INVENTION: Sarcoma-Associated Herpesvirus, DNA
; TITLE OF INVENTION: Encoding Same And Uses Thereof
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/728,323A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 0575/52268/JPW/MSC/SKS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-278-0400
; TELEFAX: 212-391-0525
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1162 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-728-323A-2

Query Match 8.4%; Score 65; DB 2; Length 1162;
Best Local Similarity 15.5%; Pred. No. 14;
Matches 18; Conservative 35; Mismatches 63; Indels 0; Gaps 0;

QY 30 LRNEERAIDELKKQATDEKATTAETAASSDALEALADQTDALQSEEAHVVKADNAASDA 89
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 776 LEDQEQEELEEEQEEQEEQEEQEEQEEQEEQEEQEEQEEQEEQEEQEEQEEQEEQEEQ 835
QY 90 LEALADQTDALQSEEAHVVSQSDNAASDAWEKAATPIALDVKKTKDPKPVVKKKEEQ 145
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 836 EQLEEQEEVEEQEEVEEQEEQEEQEEQEEQEEQEEQEEQEEQEEQEEQEEQEEQEEQEE 891

RESULT 33
US-08-216-894-2
; Sequence 2, Application US/08216894
; Patent No. 5876734
; GENERAL INFORMATION:
; APPLICANT: Kirchhoff, Louis V.
; APPLICANT: Otsu, Keiko
; TITLE OF INVENTION: POLYPEPTIDES FOR DIAGNOSING INFECTION
; TITLE OF INVENTION: WITH TRYPAOSOMA CRUZI
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
```

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 564 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-115-746-2

Query Match 8.2%; Score 64; DB 4; Length 564;  
Best Local Similarity 20.2%; Pred. No. 7.6;  
Matches 19; Conservative 20; Mismatches 55; Indels 0; Gaps 0;

QY 29 ALRNEERAIDELKKQAIEDKEATTAIEAASSDALEALADOTDLOLSEEAHVVKADNNAASD 88  
DB 367 AVTEKQRAAEATKVAEAKQKAAEAAKAVETEKQRAAEATKVAEAKQRAAEAMKVAAE 426  
QY 89 ALEALADOTDLOLSEEAHVVKQSDNNAASDAWEKAA 122  
DB 427 EKQKAAEATKVAEAKQKAAEATKVAEAKQKAA 460

RESULT 35  
US-08-216-894-8  
Sequence 8, Application US/08216894  
Patent No. 5876734  
GENERAL INFORMATION:

APPLICANT: Kirchhoff, Louis V.  
APPLICANT: Otsu, Keiko  
TITLE OF INVENTION: POLYPEPTIDES FOR DIAGNOSING INFECTION  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington, D.C.  
COUNTRY: USA  
ZIP: 20007-5109

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/216,894  
FILING DATE: 24-MAR-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 85326/102/DRLO  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136

INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 643 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-216-894-8

Query Match 8.2%; Score 64; DB 2; Length 643;  
Best Local Similarity 20.2%; Pred. No. 9;  
Matches 19; Conservative 20; Mismatches 55; Indels 0; Gaps 0;

QY 29 ALRNEERAIDELKKQAIEDKEATTAIEAASSDALEALADOTDLOLSEEAHVVKADNNAASD 88  
DB 367 AVTEKQRAAEATKVAEAKQKAAEAAKAVETEKQRAAEATKVAEAKQRAAEAMKVAAE 426

QY 89 ALEALADOTDLOLSEEAHVVKQSDNNAASDAWEKAA 122  
DB 427 EKQKAAEATKVAEAKQKAAEATKVAEAKQKAA 460

RESULT 36  
US-09-115-746-8  
Sequence 8, Application US/09115746  
Patent No. 6228601  
GENERAL INFORMATION:

APPLICANT: Kirchhoff, Louis V.  
APPLICANT: Otsu, Keiko  
TITLE OF INVENTION: POLYPEPTIDES FOR DIAGNOSING INFECTION  
TITLE OF INVENTION: WITH TRYPAOSOMA CRUZI  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington, D.C.  
COUNTRY: USA  
ZIP: 20007-5109

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/115,746  
FILING DATE:

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/216,894  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 85326/102/DRLO  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136

INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 643 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-115-746-8

Query Match 8.2%; Score 64; DB 4; Length 643;  
Best Local Similarity 20.2%; Pred. No. 9;  
Matches 19; Conservative 20; Mismatches 55; Indels 0; Gaps 0;

QY 29 ALRNEERAIDELKKQAIEDKEATTAIEAASSDALEALADOTDLOLSEEAHVVKADNNAASD 88  
DB 367 AVTEKQRAAEATKVAEAKQKAAEAAKAVETEKQRAAEATKVAEAKQRAAEAMKVAAE 426  
QY 89 ALEALADOTDLOLSEEAHVVKQSDNNAASDAWEKAA 122  
DB 427 EKQKAAEATKVAEAKQKAAEATKVAEAKQKAA 460

RESULT 37  
US-08-317-450B-15  
Sequence 15, Application US/08317450B  
Patent No. 5660982  
GENERAL INFORMATION:

APPLICANT: Tryggvason, Karl  
APPLICANT: Kallunki, Pekka  
APPLICANT: Pyke, Charles  
TITLE OF INVENTION: Laminin Chains: Diagnostic and  
TITLE OF INVENTION: Therapeutic Use  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:

ADDRESSEE: BANNER & ALLEGRETTI, LTD.  
STREET: Ten South Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC Compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/317,450B  
FILING DATE: 04-0CT-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Chao, Mark  
REGISTRATION NUMBER: 37,293  
REFERENCE/DOCKET NUMBER: 94,778  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-715-1000  
TELEFAX: 312-715-1234  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1111 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-317-450B-15

Query Match 8.2%; Score 64; DB 1; Length 1111;  
Best Local Similarity 32.3%; Pred. No. 18;  
Matches 20; Conservative 8; Mismatches 34; Indels

Qy	30	LRNEERAIDELKKQAIEDKEATTATEAASSDALEALADQTDALOSEEAUVVKADNASDA	89
		:     :   :   :     :     :     :     :     :     :	
Db	955	LRFQVDNRRKAEAEAMKRLSYISQKVSADSKTQOAEALGSAADAQRAKNAGEA	1014
Qy	90	LE 91	
Db	1015	LE 1016	

```

RESULT 38
US-08-800-593-15
; Sequence 15, Application US/08800593
; Patent No. 6143505
; GENERAL INFORMATION:
; APPLICANT: Tryggvason, Karl
; APPLICANT: Kallunki, Pekka
; APPLICANT: Pyke, Charles
; TITLE OF INVENTION: Laminin Chains: Diagnostic and
; TITLE OF INVENTION: Therapeutic Use
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
; STREET: 300 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/800,593
; FILING DATE: 18-FEB-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/317,450
; FILING DATE: 04-OCT-1994

```

```

; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Chao, Mark
; REGISTRATION NUMBER: 37,293
; REFERENCE/DOCKET NUMBER: 94,778-B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-913-0001
; TELEFAX: 312-913-0002
; INFORMATION FOR SEQ ID: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1111 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-080-593-15

```

```
Query Match      8.2%; Score 64; DB 4; Length 1111;
Best Local Similarity 32.3%; Pred. NO. 18;
Matches 20; Conservative 8; Mismatches 34; Indels 0; Gaps 0;
```

QY	30	L R N E R A I D E L K K Q A I E D K E A T T A T E A S S D A L E A L D Q T D A L O S E F A N V V K A D N A S D A	89
		: : :   :   :   :   :   :   :   :   :   :   :   :	
D b	955	L R E F L Q V D N R K A E A E M K R U S Y I S Q K V S D A S D K T Q Q A E R A L G S A A A D Q R A K N G A E A	1014
QY	90	L E	91
D b	1015	L E	1016

```

RESULT 39
US-08-317-450B-13
; Sequence 13, Application US/08317450B
; Patent No. 5660982
; GENERAL INFORMATION:
; APPLICANT: Tryggvason, Karl
; APPLICANT: Kallunki, Pekka
; APPLICANT: Pyke, Charles
; TITLE OF INVENTION: Laminin Chains: Diagnostic and
; TITLE OF INVENTION: Therapeutic Use
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BANNER & ALLEGRETTI, LTD.
; STREET: Ten South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/317,450B
; FILING DATE: 04-OCT-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Chao, Mark
; REGISTRATION NUMBER: 37,293
; REFERENCE/DOCKET NUMBER: 94,778
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-715-1000
; TELEFAX: 312-715-1234
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1193 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-317-450B-13

```

Query Match	8.2%;	Score 64;	DB 1;	Length 1193;
Best Local Similarity	32.3%;	Pred. No. 20;		

**Matches** 20; **Conservative** 8; **Mismatches** 34; **Indels** 0; **Gaps** 0;

Qy	30	LRNEERAI	DELKQAT	EDKEATT	ATEAASSD	ALEALAD	QDTAL	QSEFAN	VVKAD	NRASDA	89	
			:	:	:	:	:	:	:	:	:	
			:	:	:	:	:	:	:	:	:	
Db	955	LFRED	QVDR	NRKAE	AEEMK	RUSYIS	QKVS	DAOKT	QQAER	ALGSAAD	QRAKNGAGEA	1014
			:	:	:	:	:	:	:	:	:	
			:	:	:	:	:	:	:	:	:	
Qy	90	LE	91									
			:	:	:	:	:	:	:	:	:	
Db	1015	LE	1016									
			:	:	:	:	:	:	:	:	:	
			:	:	:	:	:	:	:	:	:	

**RESULT 40**

```

US-08-800-593-13
; Sequence 13, Application US/08800593
; Patent No. 6143505
; GENERAL INFORMATION:
; APPLICANT: Tryggvason, Karl
; APPLICANT: Kallunki, Pekka
; APPLICANT: Pyke, Charles
; TITLE OF INVENTION: Laminin Chains: Diagnostic and
; TITLE OF INVENTION: Therapeutic Use
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
; STREET: 300 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/800,593

```

APPLICATION NUMBER: US/08/800.593

FILING DATE: 18-FEB-1997

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

**APPLICATION NUMBER:**

FILING DATE: 04-OCT-  
CLASSIFICATION: 125

CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION

NAME: Chao, Mark

REGISTRATION NUMBER:

REFERENCE/DOCKET NUMB

TELECOMMUNICATION INFORMATION

TELEPHONE: 312-913-0

TELEFAX: 312-913-0000  
INFORMATION FOR CEO IS NO

### FORMATION FOR SEQU ID NO

LENGTH: 1193 amino acids

TYPE: amino acid

**TOPOLOGY:** linear

MOLECULE TYPE: protein

-800-593-13

11

Only Match

Local similarity 32  
 ches 20: Conservativ

ATTN: TROTTING, JACOB TROTTING

30 LRNEERAIDELKKQAIED

—  
—  
• •  
—  
—  
• •  
• •  
—  
—

955 LREFDLQVDNRKAEAEA

1  
 2  
 3  
 4  
 5  
 6  
 7  
 8  
 9  
 10  
 11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 100  
 101  
 102  
 103  
 104  
 105  
 106  
 107  
 108  
 109  
 110  
 111  
 112  
 113  
 114  
 115  
 116  
 117  
 118  
 119  
 120  
 121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161  
 162  
 163  
 164  
 165  
 166  
 167  
 168  
 169  
 170  
 171  
 172  
 173  
 174  
 175  
 176  
 177  
 178  
 179  
 180  
 181  
 182  
 183  
 184  
 185  
 186  
 187  
 188  
 189  
 190  
 191  
 192  
 193  
 194  
 195  
 196  
 197  
 198  
 199  
 200  
 201  
 202  
 203  
 204  
 205  
 206  
 207  
 208  
 209  
 210  
 211  
 212  
 213  
 214  
 215  
 216  
 217  
 218  
 219  
 220  
 221  
 222  
 223  
 224  
 225  
 226  
 227  
 228  
 229  
 230  
 231  
 232  
 233  
 234  
 235  
 236  
 237  
 238  
 239  
 240  
 241  
 242  
 243  
 244  
 245  
 246  
 247  
 248  
 249  
 250  
 251  
 252  
 253  
 254  
 255  
 256  
 257  
 258  
 259  
 260  
 261  
 262  
 263  
 264  
 265  
 266  
 267  
 268  
 269  
 270  
 271  
 272  
 273  
 274  
 275  
 276  
 277  
 278  
 279  
 280  
 281  
 282  
 283  
 284  
 285  
 286  
 287  
 288  
 289  
 290  
 291  
 292  
 293  
 294  
 295  
 296  
 297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312  
 313  
 314  
 315  
 316  
 317  
 318  
 319  
 320  
 321  
 322  
 323  
 324  
 325  
 326  
 327  
 328  
 329  
 330  
 331  
 332  
 333  
 334  
 335  
 336  
 337  
 338  
 339  
 340  
 341  
 342  
 343  
 344  
 345  
 346  
 347  
 348  
 349  
 350  
 351  
 352  
 353  
 354  
 355  
 356  
 357  
 358  
 359  
 360  
 361  
 362  
 363  
 364  
 365  
 366  
 367  
 368  
 369  
 370  
 371  
 372  
 373  
 374  
 375  
 376  
 377  
 378  
 379  
 380  
 381  
 382  
 383  
 384  
 385  
 386  
 387  
 388  
 389  
 390  
 391  
 392  
 393  
 394  
 395  
 396  
 397  
 398  
 399  
 400  
 401  
 402  
 403  
 404  
 405  
 406  
 407  
 408  
 409  
 410  
 411  
 412  
 413  
 414  
 415  
 416  
 417  
 418  
 419  
 420  
 421  
 422  
 423  
 424  
 425  
 426  
 427  
 428  
 429  
 430  
 431  
 432  
 433  
 434  
 435  
 436  
 437  
 438  
 439  
 440  
 441  
 442  
 443  
 444  
 445  
 446  
 447  
 448  
 449  
 450  
 451  
 452  
 453  
 454  
 455  
 456  
 457  
 458  
 459  
 460  
 461  
 462  
 463  
 464  
 465  
 466  
 467  
 468  
 469  
 470  
 471  
 472  
 473  
 474  
 475  
 476  
 477  
 478  
 479  
 480  
 481  
 482  
 483  
 484  
 485  
 486  
 487  
 488  
 489  
 490  
 491  
 492  
 493  
 494  
 495  
 496  
 497  
 498  
 499  
 500  
 501  
 502  
 503  
 504  
 505  
 506  
 507  
 508  
 509  
 510  
 511  
 512  
 513  
 514  
 515  
 516  
 517  
 518  
 519  
 520  
 521  
 522  
 523  
 524  
 525

16 37 06

1015 15 1016  
11

NOT EN COT

h completed: October 13

```
time : 27.6578 secs
```

THIS PAGE BLANK (USPTO)



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 13, 2002, 02:12:56 : Search time 1886.74 Seconds  
(without alignments)  
1137.418 Million cell updates/sec

Title: US-09-847-539A-6

Perfect score: 777

Sequence: 1 VDSPIEQRIIPNGTTLNL.....KKEQRQNVNLTPTTGESNP 159

Scoring table: BLOSUM62

Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh  
-Q/cgcn2\_1/USPTO.spool/US09847539/runat\_10102002\_092548\_3435/app\_query.fasta\_1.526  
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=40 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09847539 -CGCN\_1.1\_763 -runat\_10102002\_092548\_3435 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOPOP=6 -FGAPEXT=7  
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

EST:\*

1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_estl:\*  
10: gb\_est2:\*  
11: gb\_htc:\*  
12: gb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_inv:\*  
15: em\_gss\_pln:\*  
16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	108	13.9	442	9	AU180004
c 2	108	13.9	570	9	AU179963

C	3	104	13.4	548	9	AU177381	AU177381
C	4	99	12.7	344	9	AU180137	AU180137
C	5	99	12.7	615	9	AI057981	SMOVL3CAN
C	6	98	12.6	475	10	BM029212	BM029212
C	7	97	12.5	475	9	AU179732	IPSKA0005
C	8	96	12.4	380	9	AU179899	AU179732
C	9	88	11.2	567	12	AZ895651	AZ895651
C	10	87	11.2	895	12	AZ534176	ENTBX38TR
C	11	86	11.1	394	9	AU179547	AU179547
C	12	86	11.1	833	12	CNS06072	AL407733
C	13	85	10.9	1097	12	CNS06NLI	T7 end of
C	14	84	10.8	577	12	BH363656	AL406907
C	15	84	10.8	745	12	CNS06OVL	BH363656
C	16	83	10.7	994	12	CNS07EBO	CH230-203
C	17	81	10.4	453	9	AU179269	AL408583
C	18	81	10.4	857	10	BG837033	T7 end of
C	19	80	10.3	308	9	AU171443	AL41562
C	20	80	10.3	744	9	AU169406	T7 end of
C	21	80	10.3	1123	11	AK005846	AL41562
C	22	80	10.3	1158	11	AK014987	Mus muscu
C	23	79	10.2	556	12	TA249H11Q	AK014987
C	24	79	10.2	797	12	AQ875889	AL482135
C	25	79	10.2	850	12	AZ168111	T. brucei
C	26	79	10.2	928	12	CNS03T32	AQ875889
C	27	79	10.2	1076	12	CNS06MOG	VI30F6 mT
C	28	78	10.0	338	12	FR0042547	AL168111
C	29	78	10.0	494	12	AQ875428	SP_0103_B
C	30	78	10.0	498	10	B1863089	AL259319
C	31	78	10.0	623	12	AQ504321	Tetraodon
C	32	78	10.0	753	9	AU003803	AL45734
C	33	78	10.0	755	9	AJ395963	T3 end of
C	34	78	10.0	968	12	CNS07CXJ	AL130039
C	35	77	9.9	394	12	AQ911873	Fugu rubr
C	36	77	9.9	425	9	AA784607	AL130039
C	37	77	9.9	588	10	BJ039902	VI25C3 mT
C	38	77	9.9	663	10	BF483682	BI863069
C	39	77	9.9	675	9	AU003869	rm16b06.y
C	40	77	9.9	723	12	AZ948693	AL427196
C	41	77	9.9	750	10	BG778500	clone BAO
C	42	77	9.9	785	10	BM405132	Tetraodon
C	43	77	9.9	790	12	CNS07270	clone BAO
C	44	77	9.9	935	12	CNS033D4	clone BAO
C	45	77	9.9	1061	12	CNS0738M	clone BAO

ALIGNMENTS

RESULT 1  
AU180004/c

LOCUS

DEFINITION

442 bp mRNA linear EST 21-MAR-2001

Medaka clone Olcl8.01d similar to gb|AB025967| Oryzias latipes mRNA

for chorionogenin Hminor, complete cds, mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

1 (bases 1 to 442)

Naruse, K., Mitani, H. and Tanaka, M.

Medaka EST Project in University of Tokyo (2001)

Unpublished (2001)

Contact: Kiyoshi Naruse

Department of Biological Sciences

Graduate School of Science, University of Tokyo

Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, Japan

Tel: 81-3-5841-4443

Fax: 81-3-5841-4410

Email: naruse@biol.s.u-tokyo.ac.jp



Tel: 81-3-5841-4443  
Fax: 81-3-5841-4410  
Email: naruse@iol.s.u-tokyo.ac.jp  
This clone was isolated from Medaka liver cDNA library (OLE)  
Location/Qualifiers

```

FEATURES
  source          "This clone was isolated from Medaka liver cDNA library (OLE)."
  location        "Location/Qualifiers"
  accession       "1344"
  organism        "Oryzias latipes"
  strain          "HN1"
  db_xref        "taxon:8090"
  clone_name      "clone_OLE9.1le"
  clone_lib       "Medaka liver cDNA library (OLE) from HN1"
  tissue_type     "liver"

```

BASE COUNT	96 a	132 c	41 g	74 t	1 others
ORIGIN					
/dev_stage=					

Alignment Scores:	
Pred. No.:	0.00981
Score:	99.00
Percent Similarity:	43.01%
Best Local Similarity:	31.18%
Query Match:	12.74%
DB:	9
Length:	344
Matches:	29
Conservative:	11
Mismatches:	53
Indels:	0
Gaps:	0

US-09-847-539A-6 (1-159) x AU180137 (1-344)

Qy	24	AlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaIleAspGluLeuLysLysGln	43
		::       ::	
Db	330	GCTGAGGATNTGAGGTTTTGAGGGTACTGAGCGTCTGAGGATCTCTGAGGATGTTGAG	271
Qy	44	AlaIleGluAspLysGluAlaThrThrAlaIleGluAlaAlaSerSerAspAlaLeuGlu	63
Db	270	GCTGCTGAGGATTTCTGAGGCTTTGAAGGATACTGAGCGTCTGAGGATCTCTGAGGATGTT	211
Qy	64	AlaLeuAlaAspGlnThrAspAlaLeuGlnInsrGluGluAlaAlaValValLysAlaAsp	83
		::  ::::  ::   ::       ::	
Db	210	GAGCGCTGAGGATCTGAGGTTTGAAGGATCTGAGCGTCTCTGAGGATCTCTGAGGAT	151
Qy	84	AsnAlaAlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGlu	103
		::   ::	
Db	150	ACTGAGGCTGATGAGGCTGCTGAGGTTTGAAGGATCTCTGAGGCTCTCTGAGGATCTCTGAG	91

LOCUS	615 bp	linear	EST 20-JUL-1998
DEFINITION	Onchocerca volvulus infective larva cDNA		
	SAW94WL-OvL3	Onchocerca volvulus cDNA clone SWOVL3CAN21G06 5', mRNA sequence.	

ORGANISM: *Onchocerca volvulus*  
Eukaryota: Metazoa: Ne

REFERENCE AUTHORS TITLE JOURNAL COMMENT
Onchocercidae: Onchocerca. 1 (bases 1 to 615) Williams, S.A., Lu, W., Lizotte-Waniewski, M. and Laney, S.J. Genes expressed in infective third stage larvae of Onchocerca volvulus Unpublished (1995) Contact: Steven A. Williams Molecular Parasitology Smith College Department of Biological Sciences Department of Biological Sciences, Clark Science Center, Smith College, Northampton, MA, 01063, USA

College, Northampton, MA, 01063, USA  
Tel: 4135853826  
Fax: 4135853786  
Email: [genomesmith.edu](mailto:genomesmith.edu)  
Seq primer: phaluescript SK.

FEATURES  
source Location/Qualifiers  
1. .615  
/organism="Onchocerca volvulus"  
/strain="Sierra Leone"  
/db\_xref="taxon:6282"  
/clone="SNOVL3CAN21G06"  
/clone\_lib="Onchocerca volvulus infective larva cDNA  
(SAW94WL-Ovi3)"  
/lab\_host="XLI-Blue MRP"  
/note="Vector: Lambda UniZap XR; Site1: EcoR I; Site2:  
Xho I; Cutaneous filarial nematode parasite of humans"  
RNA was prepared from third stage infective larvae of  
Onchocerca volvulus isolated from mosquitoes 10 days after  
infection and converted to double stranded cDNA using  
reverse transcriptase and oligo(dT) followed by RNase H  
and DNapol I. The library had 1.8 x 10E5 independent  
recombinants and average insert size was 900 base pairs.  
The library was constructed by Wenhong Lu. The library is  
available from Dr. S.A. Williams, email genome@smith.edu."

BASE COUNT 74 a 186 c 173 t 6 others  
ORIGIN

Alignment Scores:  
Pred. No.: 0.0202 Length: 615  
Score: 99.00 Matches: 26  
Percent Similarity: 53.16% Conservative: 16  
Best Local Similarity: 32.91% Mismatches: 37  
Query Match: 12.74% Indels: 0  
DB: 9 Gaps: 0

US-09-847-539A-6 (1-159) x A1057981 (1-615)

QY	50	AlaThrThrAlaTleGluAlaAaSerSerAspAlaLeuGluAlaLeuAlaAaspGlnThr	69
DB	471	GCAGAGTCCTCAAAAACGGCGCGCCACGAGTCCCGGCGGCGGAAAACGTCAGAAACG	412
QY	70	AspAlaLeuGlnSerGluGluAlaAlaValValLysAlaAaspAsnAlaAaSerAspAla	89
	:::		
DB	411	AATGCTCGACGGTCACAAACAATCAGCGCCACGCTCGNTCCACCGCGGCACGAAACGC	352
QY	90	LeuGluAlaLeuAlaAaspGlnThrAspAlaLeuGlnSerGluGluAlaGluValValGln	109
DB	351	TCAGAGCGCCGCCACTTCAGCACGAGATCGCGTGCGCTCAAAAGAGCGCAAAATCATCA	292
QY	110	SerAspAsnAlaAaSerAspAlaTlpGluLysAlaAlaThrProfilAlaLeuAasp	128
DB	291	GAACGAAACGCATCATCAAGTCGCGGTGTCAGACATTCCTCGGCACGCGCGGACGAA	235

RESULT 6  
BM029212 475 bp mRNA linear EST 05-NOV-2001  
LOCUS BM029212 Skin cDNA library Ictalurus punctatus cDNA 5', mRNA  
DEFINITION sequence.  
ACCESSION BM029212  
VERSION BM029212.1 GI:16742782  
KEYWORDS EST.  
SOURCE channel catfish.  
ORGANISM Ictalurus punctatus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Siluriformes;  
Ictaluridae; Ictalurus.  
REFERENCE 1 (bases 1 to 475)  
AUTHORS Karst, A., Cao, D., Li, P., Ju, Z., Kocabas, A., Feng, J., Patterson, A.,  
Mickett, K.D. and Liu, Z.  
TITLE Transcriptome analysis of channel catfish (Ictalurus punctatus):  
Initial analysis of gene expression and microsatellite-containing  
cDNAs in the skin  
JOURNAL Unpublished (2001)  
COMMENT Contact: Liu, ZJ  
The Fish Molecular Genetics and Biotechnology Laboratory,  
Department of Fisheries and Allied Aquacultures and Program of Cell  
and Molecular Biosciences  
Auburn University

203 Swingle Hall, Auburn University, Auburn, AL 36849, USA  
Tel: 334 844 4054  
Fax: 334 844 9208  
Email: zlin@acesag.auburn.edu  
Seq primer: M13 Reverse.  
  
FEATURES  
    source  
        Location/Qualifiers  
            1..475  
            /organism="Ictalurus punctatus"  
            /db\_xref="taxon:7998"  
            /clone\_lib="skin cDNA library"  
            /note="Organ: Skin; Vector: pSport1; Site\_1: NotI; Site\_2:  
            SalI"

BASE COUNT     229 a     77 c     154 g     15 t

ORIGIN

Alignment Scores:

Pred. No.:	Score:	Length:
Percent Similarity:	98-100	Matches: 475
Best Local Similarity:	48-39%	Conservative: 27
Query Match:	29.03%	Mismatches: 18
	12.61%	Indels: 48
	10	Gaps: 0

US-09-847-539A-6 (1-159) x BM029212 (1-475)

QY 26 GlulysLeuAlaLeuArgAsnGluGluArgAlaIleAspGluLeuLysLysGlnAlaIle 45  
||| ||| ||| ||| ||| : | : : :  
Db 163 GAACCACAGCAGCAAGAAGCAGCAGAGAGACGACGACGAAGAAAGAACGACGACGA 222  
||| ||| ||| ||| ||| : | : : :  
QY 46 GluasplysGluAlaThrThrAlaIleGluAlaAlaSerSerAspAlaLeuGluAlaLeu 65  
||||: ||| ||| ||| ||| : : : : :  
Db 223 GAGAAGCAGAGAAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAA 282  
||||: ||| ||| ||| ||| : : : : :  
QY 66 AlaaspGlnThrAspAlaLeuGlnSerGluGluAlaValvallyAlaAspAsnAla 85  
||||: ||| ||| ||| ||| ||| ||| ||| : : : : :  
Db 283 CGAAGCAGAGAAGCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGACA 342  
||| ||| ||| ||| ||| ||| ||| ||| : : : : :  
QY 86 AlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAla 105  
||| ||| ||| ||| ||| ||| ||| ||| : : : : :  
Db 343 GAACCACAGCAGNAGACA 402  
||| ||| ||| ||| ||| ||| ||| ||| : : : : :  
QY 106 GluValValGlnSerAspAsnAlaAlaSerAspAlaTrp 118  
||| ||| ||| ||| ||| ||| ||| ||| : : : : :  
Db 403 GAAGCAGCAGCAGAGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATTT 441  
||| ||| ||| ||| ||| ||| ||| ||| : : : : :

RESULT 7

AUI179732/c

LOCUS AUI179732 Medaka liver cdna library (OLE) from HNI Oryzias latipes  
DEFINITION cdna clone OLEj3.10d similar to gb|AB025967| Oryzias latipes mRNA  
for chorionogenin Hminor, complete cds, mRNA sequence.

ACCESSION AUI179732

VERSION AUI179732.1 GI:13428569

KEYWORDS EST.

SOURCE Japanese medaka.  
Oryzias latipes

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
Acanthomorphia; Acanthopterygii; Percomorpha; Atherinomorpha;  
Belontiiformes; Adrianichthyidae; Oryziinae; Oryzias.  
1 (bases 1 to 517)  
Naruse,K., Mitani,H. and Tanaka,M.  
Medaka EST Project in University of Tokyo (2001)  
Unpublished (2001)  
Contact: Kiyoshi Naruse  
Department of Biological Sciences  
Graduate School of Science, University of Tokyo  
Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, Japan  
Tel: 81-3-5841-4443  
Fax: 81-3-5841-4410  
Email: naruse@biol.s.u-tokyo.ac.jp  
This clone was isolated from Medaka liver cdna library (OLE).  
Location/Qualifiers

FEATURES





```

KEYWORDS EST.
SOURCE Japanese medaka.
ORGANISM Oryzias latipes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
Belontiiformes; Adrianichthyidae; Oryziatidae; Oryzias.
REFERENCE 1 (bases 1 to 394)
AUTHORS Naruse, K., Mitani, H. and Tanaka, M.
TITLE Medaka EST Project in University of Tokyo (2001)
JOURNAL Unpublished (2001)
COMMENT Contact: Kiyoshi Naruse
Department of Biological Sciences
Graduate School of Science, University of Tokyo
Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, Japan
Tel: 81-3-5841-4443
Fax: 81-3-5841-4410
Email: naruse@biol.s.u-tokyo.ac.jp
This clone was isolated from Medaka liver cDNA library (OLE).
FEATURES
source
1..394
/organism="Oryzias latipes"
/strain="HNI"
/db_xref="taxon:8090"
/clone_lib="Medaka liver cDNA library (OLE) from HNI"
/tissue_type="liver"
/dev_stage="adult"
BASE COUNT 116 a 143 c 52 g 79 t 4 others
ORIGIN
Alignment Scores:
Pred. No.: 0.487 Length: 394
Score: 86.00 Matches: 25
Percent Similarity: 42.86% Conservative: 11
Best Local Similarity: 29.76% Mismatches: 48
Query Match: 11.07% Indels: 0
DB: 9 Gaps: 0
US-09-847-539a-6 (1-159) x AU179547 (1-394)
QY 33 GluGluArgAlaIleAspGluLeuLysGlnAlaIleGluAspLysGluAlaThrThr 52
Db 252 GATGAGGCTGCTAGGATTCCTGAGGATCTAGGCGGCTGAGGATTCCTGAGGTTTGAAG 193
QY 53 AlaIleGluAlaSerSerAspAlaLeuGluAlaLeuAlaSerAspAlaLeu 72
Db 192 GGTACTGAGGCTGCTGAGGATTCCTNAGGATGNTGAGNCTGCTGAGGATTCCTGAGGTTTGT 133
QY 73 GlnSerGluGluAlaAlaValValLysAlaAspAsnAlaAlaSerAspAlaLeuGluAla 92
Db 132 AGGATAGTACTGAGGCTGCTGAGGATTCCTGAGGATGTTGAGGCTGCTGAGGATTCCTGAGGTT 73
QY 93 LeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaGluValValGlnSerAspAsn 112
Db 72 TTGAAGGATACCTGAGGCTGCTGAGGATTCCTGAGGATTCCTGAGGATTCCTGAGGCTGCTGAG 13
QY 113 AlaAlaSerAsp 116
Db 12 GTTTTGAAGGAT 1
RESULT 12
CNS0607Z/c 833 bp DNA linear GSS 04-JUL-2001
LOCUS T7 end of clone AV0AA004C03 of library AV0AA from strain CBS 379 of
DEFINITION Saccharomyces exiguus, genomic survey sequence.
ACCESSION AL407733
VERSION AL407733.1 GI:12173672
KEYWORDS GSS.
SOURCE Saccharomyces exiguus.
ORGANISM Saccharomyces exiguus
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

```

```

REFERENCE 1 (bases 1 to 833)
AUTHORS Bon, E., Neuveglise, C., Lepingle, A., Wincker, P., Artiguenave, F.,
Gaillardin, C. and Casaregola, S.
TITLE Genomic exploration of the hemiascomycetous yeasts: 6.
JOURNAL Saccharomyces exiguus
MEDLINE FEMS Lett. 487 (1), 42-46 (2000)
REFERENCE 20584716
REFERENCE 2 (bases 1 to 833)
AUTHORS Souciet, J.-L., Aigle, M., Artiguenave, F., Blandin, G.,
Bolotin-Fukuhara, M., Bon, E., Brottier, P., Casaregola, S.,
de-Montigny, J., Dujon, B., Durrens, P., Lepingle, A., Llorente, B.,
Maupertuy, A., Neuveglise, C., Ozier-Kalogeropoulos, O., Potier, S.,
Saurin, W., Tekala, F., Toffano-Nioche, C., Wesolowski-Louvel, M.,
Wincker, P. and Weissbach, J.
TITLE Genomic exploration of the hemiascomycetous yeasts: 1. A set of
yeast species for molecular evolution studies
JOURNAL FEMS Lett. 487 (1), 3-12 (2000)
MEDLINE 20584711
REFERENCE 3 (bases 1 to 833)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,
2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT This GSS is part of a random genomic sequencing program of thirteen
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
5 kb were prepared and both extremities were sequenced. See
keywords for description of this sequence and for the sequence of
the other extremity of this insert.
FEATURES
source
1..833
/organism="Saccharomyces exiguus"
/strain="CBS 379"
/db_xref="taxon:34358"
/clone_lib="AV0AA004C03"
/clone_lib="AV0AA"
/clone_end="77"
<31..>258
/feature="similar to Saccharomyces cerevisiae ORF YBR067c [
TIPI1 ; esterase ]"
/evidence="not_experimental"
misc_feature
<31..>255
/feature="similar to Saccharomyces cerevisiae ORF YOR010c [
TIR2 ; cold shock induced protein ]"
/evidence="not_experimental"
misc_feature
<31..>225
/feature="similar to Saccharomyces cerevisiae ORF YER011w [
TIR1 ; cold-shock induced protein of the Tirip, Tirip
family ]"
/evidence="not_experimental"
misc_feature
<34..>243
/feature="similar to Saccharomyces cerevisiae ORF YOR009w [
similarity to Tirip and Tir2p ]"
/evidence="not_experimental"
BASE COUNT 165 a 191 c 128 g 332 t 17 others
ORIGIN
Alignment Scores:
Pred. No.: 1.24 Length: 833
Score: 86.00 Matches: 21
Percent Similarity: 48.81% Conservative: 20
Best Local Similarity: 25.00% Mismatches: 43
Query Match: 11.07% Indels: 0
DB: 12 Gaps: 0
US-09-847-539a-6 (1-159) x CNS0607Z (1-833)
QY 33 GluGluArgAlaIleAspGluLeuLysGlnAlaIleGluAspLysGluAlaThrThr 52

```







Genoscope.  
Direct Submission  
Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,  
2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :  
seqref@genoscope.cns.fr - Web :  
This GSS is part of a random genomic sequencing program of thirteen  
yeast species: *Saccharomyces bayanus* var. *uvarum*, *Saccharomyces*  
*exiguus*, *Saccharomyces servazzii*, *Zygosaccharomyces rouxii*,  
*Saccharomyces kluyveri*, *Kluyveromyces thermotolerans*, *Kluyveromyces*  
*lactis* var. *lactis*, *Kluyveromyces marxianus* var. *marxianus*, *Pichia*  
*angusta*, *Debaryomyces hansenii* var. *hansenii*, *Pichia sorbitophila*,  
*Candida tropicalis* and *Yarrowia lipolytica*. Genomic inserts of 3 to  
5 kb were prepared and both extremities were sequenced. See  
keywords for description of this sequence and for the sequence of  
the other extremity of this insert.

[illegible]

RESULT 17	AUI79269/c
LOCUS	AUI79269 Medaka liver cDNA library (Ole) EST 21-MAR-2001
DEFINITION	cDNA clone Ole06_03e similar to gb AB025967  Oryzias latipes mRNA for chorionenin Hminor, complete cds, mRNA sequence.
ACCESSION	AUI79269
VERSION	AUI79269.1 GI:13428106
KEYWORDS	EST.
SOURCE	Japanese medaka.
ORGANISM	Oryzias latipes
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopercyiformes; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percormorpha; Atherinomorpha; Belontiiformes; Adrianichthyidae; Orziilinae; Oryziinae; Oryzias. 1 (bases 1 to 453)
REFERENCE	Naruse,K., Mitani,H. and Tanaka,M.

TITLE Medaka EST Project in University of Tokyo (2001)  
JOURNAL Unpublished (2001)  
COMMENT Contact: Kiyoshi Naruse  
Department of Biological Sciences  
Graduate School of Science, University of Tokyo  
Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, Japan  
Tel: 81-3-5841-4443  
Fax: 81-3-5841-4410  
Email: naruse@biol.s.u-tokyo.ac.jp  
This clone was isolated from Medaka liver cDNA library (OLE).

FEATURES	Source
Location/Qualifiers	1..453
Organism="Oryzias latipes"	
/strain="HNI"	
/db_xref="taxon:8090"	
/clone="Ole06.03e"	
/clone_lib="Medaka liver cDNA library (Ole) from HNI"	
/tissue_type="liver"	
/dev_stage="adult"	
137 a	159 c 59 g 98 t
BASE COUNT	
ORIGIN	

Alignment Scores:			
pred. No.:	2.44	Length:	453
Score:	81.00	Matches:	24
Percent Similarity:	44.59%	Conservative:	9
Best Local Similarity:	32.43%	Mismatches:	41
Query Match:	10.42%	Indels:	0
DB:	9	Gaps:	0
US-09-847-539A-6 (1-159) x AUL179269 (1-453)			
QY	33	GUUGUArgAlaIleAspGLeuLysGlnAlaIleGluAspLysGluAlaThrThr	52
Db	236	GATGAGGCTGCTTGAGGATCTTGAGGACTCTGAGCGGCTGAGGATTTCTGAGGTTTGAAG	17
QY	53	AlaIleGluAlaAlaSerSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeu	72
Db	176	GGTACTGAGGCTGCTGAGGATTTCTGAGGATGTTTGAGGCTGCTGAGGATTTCTGAGGTTTG	11
QY	73	GlnSerGluGluAlaAlaValIlysAlaAspAsnAlaAlaSerAspAlaLeuGluAla	92
Db	116	AAGGACTATGAGGCTGCTGAGGATTTCTGAGGATGTTTGAGGCTGCTGAGGATTTCTGAGGTT	57
QY	93	LeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaGlu	106
Db	56	TTGAGGAGTACTGAGGCTGCTGAGGATTTCTGAGGATTTCTGAGGATTTCTGAGGTT	15

LOCUS	857 bp	mrna	linear	EST 25-MAY-200
DEFINITION	zmo8_10h09_A	Fusarium_graminearum_inoculated_corn_ear	zmo8_10h09	zmo8_10h09
		CDNA clone zmo8_10h09	zmo8_10h09	zmo8_10h09
		sequence.		

ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS
BG837033	1	GI:14203356	EST.	zea mays.	1	Harris, L.J., Balcerzak, M., Allard, S., Saparno, A., Couroux, P., De Moors, A., Hattori, J.I., Ouellet, T., Robert, L.S., Singh, J.A., Sprott, D. and Tinker, N.A.
				zea mays.	1 (bases 1 to 857)	
				zea mays		
				Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Zea.		

TITLE	JOURNAL	COMMENT
<p>U. and Linker, N.N.  Expressed Sequence Tags from Developing Maize Kernels Six Days after Silk Channel Inoculation with <i>Fusarium graminearum</i>  Unpublished (2001)  Contact: Harris, Linda J.  Eastern Cereal and Oilseed Research Centre  Agriculture and Agri-food Canada  Bldg. 21, Central Experimental Farm, Ottawa, Ontario, K1A 0C6,</p>		



BASE COUNT	116 a	178 c	238 g	212 t
ORIGIN	/tissue_type="brain" /dev_stage="adult"			
Alignment Scores:	6.03	Length:	744	
Pred. No.:	80.00	Matches:	21	
Score:	57.53%	Conservative:	21	
Percent Similarity:	28.77%	Mismatches:	31	
Best Local Similarity:	10.30%	Indels:	0	
Query Match:	9	Gaps:	0	
DB:				
US-09-847-539a-6 (1-159) x AU169406 (1-744)				
Qy	50	AlaThrThrAlaIleGluAlaAlaSerSerAspAlaLeuGluAlaLeuAlaAaspGlnThr	69	
Db	279	GCAACCCAGCCGCTGACAGCGTGCAGCCAGCAGCTCTCAGTTAGAGCAGTCGACCC	220	
Qy	70	AspAlaLeuGlnSerGluGluAlaAlaValValLysAlaAspAlaAlaSerAspAla	89	
Db	219	AGCCGCTGACAGCGTGCAGCCAGCGCAGTCACAGTCAGCAGCGTCGACCGCGCA	160	
Qy	90	LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaGluValValGln	109	
Db	159	GTCTGACGAGCTGCAGCCAGCAGCAGCTCTCAGTTAGAGCAGTCGACCGCAGTCAGA	100	
Qy	110	SerAspAsnAlaAlaSerAspAlaTrpGluLysAlaAla	122	
Db	99	GCAGCTGCAGCCAGCGCAGCTCTCAGTCAGTCAGCAGCAGCTGCA	61	
RESULT 21				
LOCUS	AK005846	1123 bp	mrna	linear
DEFINITION	Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:1700010M08:hypothetical protein, full insert sequence.			
ACCESSION	AK005846	1	GI:12838633	
VERSION	AK005846	1	GI:12838633	
KEYWORDS	HTC; CAP trapper			
SOURCE	Mus musculus (Strain:C57BL/6J) adult male testis cDNA to mRNA, clone_lib:RIKEN full-length enriched mouse cDNA library			
ORGANISM	Mus musculus			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
AUTHORS	1 (sites)			
TITLE	Carninci, P. and Hayashizaki, Y.			
JOURNAL	High-efficiency full-length cDNA cloning			
MEDLINE	Meth. Enzymol. 303, 19-44 (1999)			
PUBMED	99279253			
REFERENCE	2 (sites)			
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.			
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes			
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)			
MEDLINE	20493374			
PUBMED	11042159			
REFERENCE	3 (sites)			
AUTHORS	Shibata, K., Itoh, M., Alizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Wachihi, M., Itoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.			
TITLE	RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer			
JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)			
MEDLINE	20530913			
PUBMED	11076861			
REFERENCE	4 (sites)			
AUTHORS	The RIKEN Genome Exploration Research Group Phase II Team and the PANTOM Consortium.			
TITLE	Functional annotation of a full-length mouse cDNA collection			
JOURNAL	Nature 409, 685-690 (2001)			
REFERENCE	5 (bases 1 to 1123)			
AUTHORS	Adachi, J., Alizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A., Arakawa, T., Baldarelli, R., Bono, H., Brownstein, M., Bult, C., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hill, D., Hiramoto, K., Hiraoka, T., Hori, F., Hume, D., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Quackenbush, J., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Schriml, L., Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T., Tejima, Y., Toya, T., Yamamura, T., Yamanaka, I., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.			
TITLE	Direct Submission			
JOURNAL	Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)			
COMMENT	Please visit our web site (http://genome.gsc.riken.go.jp/) for further details. cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN, Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAGAGAGATCCAGAGCTCTTTTTTTTTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGAGCGCGCAATTAATTCGAGTTAATTAATTCCTCCCTCCCTCC 3'] cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI. Host: SOLR.			
FEATURES	Location/Qualifiers			
source	1. .1123			
	/organism="Mus musculus"			
	/strain="C57BL/6J"			
	/db_xref="MGI:1897520"			
	/db_xref="taxon:10090"			
	/clone="1700010M08"			
	/sex="male"			
	/tissue_type="testis"			
	/clone_lib="RIKEN full-length enriched mouse cDNA library"			
	/dev_stage="adult"			
	160. .984			
	/note="evidence:NAS"			
	hypothetical protein			
	putative"			
	/codon_start=1			
	/protein_id="BAB24271.1"			
	/db_xref="GI:12838634"			
	/translation="MDVSGSQIHKRIQLFSCFCTRNQKKLHKHSHEVEKKLCRE NKALRDENLRKDNKFLWGNKALGRNKTFRMDNQFIRERNRLQQNLKRYKR LNSNPKLSGEYNALSTGKSPWQONRAEQAITLRQOEKAFONKALHEEIKSL CETKALQHOERALMEERKALMRDGVAAELAPALTKEGALEMEERALKKECALREE NKALREHICALODEEVALQEEAKILENNNLGGKITNNLPCKTQNDPKKCGPRM"			
	1105. .1110			
	/note="putative"			
	1123			
	/note="putative"			
	394 a 233 c 295 g 201 t			
BASE COUNT	394 a	233 c	295 g	201 t
ORIGIN				
Alignment Scores:				



```

Query Match: 10.30% Indels: 0
DB: 11 Gaps: 0
US-09-847-539A-6 (1-159) x AK014987 (1-1158)

QY 33 GluGluArgAlaIleAspGluLeuLysGlnAlaIleGluAspLysGluAlaThrThr 52
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 674 GAGGACCAAGGCTCTGACACACAGGAGGCTCTCAGAGGAGGAGGAGGCTCTG 733
QY 53 AlaIleGluAlaAlaSerSerAspAlaLeuGluAlaLeuAlaSerGlnThrAspAlaLeu 72
: : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 734 ATGAGGAGTGGAGTGGCGGCGGAGCTGGCGGAGGCTCTACCAAGGAGGCGGCTGCTCG 793
QY 73 GlnSerGluGluAlaValValValLysAlaAspAsnAlaAlaSerAspAlaLeuGluAla 92
: : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 794 GAGATGGAAGACAGGCTCTGTGTGAGGAGGAGGAGGCTCTCCGGGAGGAGCAAGCA 853
QY 93 LeuAlaAspGlnThrAspAlaLeuGlnSerGluGlu 104
||| : : : : : : : : : : : : : : : : : : : : :
Db 854 CTCAGAGAAGAGCATGGGCTCTCCAAGATGAAGAA 889

RESULT 23
TA249H11Q/c TA249H11Q 556 bp DNA linear GSS 13-DEC-2000
LOCUS T. brucei sheared genomic DNA clone 249h11, reverse sequence,
DEFINITION genomic survey sequence.
ACCESSION AL482135
VERSION AL482135.1 GI:11848140
KEYWORDS GSS
SOURCE Trypanosoma brucei.
ORGANISM Trypanosoma brucei.
REFERENCE 1 (bases 1 to 556)
AUTHORS Hall, N., Bowman, S., Lennard, N.J., Doggett, J., Atkin, R.,
Chillingworth, C., Ormond, D., Harris, B., El-Sayed, N., Hou, L.,
Melville, S.E., Rajandream, M.A. and Barrell, B.G.
DIRECT SUBMISSION
Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing
project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,
Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and
nhls@sanger.ac.uk
COMMENT Constructed at the Institute for Genomic Research (TIGR),
Rockville, MD. Genomic DNA isolated from a cloned population of
Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared
to give a tight size distribution (
4 kb). The v + i method used for the library construction is
described in detail in Smith, H. and Venter, J.C. (Making small
Insert libraries for whole genome shotgun sequencing projects. In
Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barrell, Oxford University Press, 1999).
Email: nelsayed@tigr.org
Details of T. brucei sequencing at the Sanger Centre are available
at http://www.sanger.ac.uk/projects/T_brucei/.

FEATURES
source
1..556
Location/Qualifiers
/organism="Trypanosoma brucei"
/strain="TREU927"
/db_xref="taxon:5691"
/clone="249h11"
BASE COUNT 88 a 186 c 136 g 146 t
ORIGIN

Alignment Scores:
Pred. No.: 5.6 Length: 556
Score: 79.00 Matches: 24
Percent Similarity: 42.50% Conservative: 27
Best Local Similarity: 20.00% Mismatches: 69
Query Match: 10.17% Indels: 0
DB: 12 Gaps: 0

US-09-847-539A-6 (1-159) x TA249H11Q (1-556)

```

```

QY 38 AspGluLeuLysLysGlnAlaIleGluAspLysGluAlaThrThrAlaIleGluAlaAla 57
: : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 504 GAACAAATCAGGCGCTCTATTCAAAATGACGACAGCGAGTGAAGACGCGGAGCAAGCG 445
QY 58 SerSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAla 77
||| : : : : : : : : : : : : : : : : : : : : :
Db 444 GTGGAGCAGCGCGGAGCGCTCTGAGGAGGCGAGAAATGACACAGATCCCGCTAAGGAT 385
QY 78 AlaValValLysAlaAspAsnAlaAlaSerAspAlaLeuGluAlaLeuAlaAspGlnThr 97
||||| : : : : : : : : : : : : : : : : : : : : :
Db 384 GCGTCCGAACCTCGGATCGGATCTCAAGATGCGAAGCGCGCTCTGGGGACCTCAGC 325
QY 98 AspAlaLeuGlnSerGluGluAlaGluValValGlnSerAspAlaAlaSerAspAla 117
: : : : : : : : : : : : : : : : : : : : :
Db 324 AGGGACATGCACCAACCAAGCGCGAGAGGAGCGCGCTCCGCGAGAAGAGGATCGC 265
QY 118 TrpGluLysAlaAlaThrProLeuAlaLeuAspValLysLysThrLysAspThrLysPro 137
: : : : : : : : : : : : : : : : : : : : :
Db 264 GAGAAGAACTCTTCGAGATTCCTGAGCCTGAAAGAGGCTTCGGAAGACCGCACAAGAG 205
QY 138 ValValLysLysGluGluArgGlnAsnValAsnThrLeuProThrThrGluGluSer 157
||| : : : : : : : : : : : : : : : : : : : : :
Db 204 GTGGCTGTGAGCTGAGAAGCAGGAGCGCGCTTTCTCTCGCTGGGAGCGAGT 145

RESULT 24
AQ875889/c AQ875889 797 bp DNA linear GSS 08-NOV-1999
LOCUS V130F6 mTn-3xHA/lacZ Insertion Library, strain Y2278 Saccharomyces
DEFINITION cerevisiae genomic 5', DNA sequence.
ACCESSION AQ875889
VERSION AQ875889.1 GI:6288133
KEYWORDS GSS.
SOURCE baker's yeast.
ORGANISM Saccharomyces cerevisiae
REFERENCE 1 (bases 1 to 797)
AUTHORS Ross-Macdonald, P., Roemer, T., Coelho, P.S.R., Agarwal, S., Kumar, A.,
Desjardes, S.A., Cheung, K.-H., Sheehan, A., Symoniat, D., Jansen, R.,
umansky, L., Heidman, M., Nelson, K., Iwasaki, H., Kanada, D., Lugo, R.,
Hager, K., Miller, P., Roeder, G.S. and Snyder, M.
Large-Scale Analysis of the Yeast Genome by Transposon Tagging and
Gene Disruption
Unpublished (1999)
Contact: Kumar A
Michael Snyder, Dept. of Mol. Cell. and Dev. Biology
Yale University
P.O. Box 208103, New Haven, CT 06520-8103, USA
Tel: 203 432 9949
Fax: 203 432 6161
Email: anuj.kumar@yale.edu
te of mTn-3xHA/lacZ insertion.
Seq primer: GCGCTCTCTCTTCTTGGAGTAC
Class: transposon-tagged.
FEATURES
source
1..797
Location/Qualifiers
/organism="Saccharomyces cerevisiae"
/strain="Y2278 - S288C background, cir(0) rho(0)"
/db_xref="taxon:4932"
/clone_lib="mtn-3xHA/lacZ Insertion Library, strain Y2278"
/lab_host="E. coli"
/note="Vector: pHS56-Sal; A yeast genomic DNA library
without 2 micron or mitochondrial DNA was prepared in
pHS56-Sal; genomic DNA was size-fractionated (DNA of
roughly 2-3 kb in length) prior to cloning. This library
was subsequently mutagenized with a mtn-3xHA/lacZ
minitransposon containing lacZ, URA3, and tet resistance."
BASE COUNT 206 a 175 c 209 g 207 t
ORIGIN

Alignment Scores:
Pred. No.: 8.76 Length: 797

```







/clone\_lib="cosmid 152F05"  
/clone="152F05BE3"

BASE COUNT 55 a 122 c 72 g 84 t 5 others  
ORIGIN

## Alignment Scores:

Pred. No.: 4.01 Length: 338  
Score: 78.00 Matches: 21  
Percent Similarity: 48.78% Conservative: 19  
Best Local Similarity: 25.61% Mismatches: 42  
Query Match: 10.04% Indels: 0  
DB: -12 Gaps: 0

US-09-847-539A-6 (1-159) x FR0042547 (1-338)

Qy 31 ArgAsnGluGluArgAlaIleAspGluLeuLysGlnAlaIleGluAspLysGluAla 50

Db 295 CGACAGATGGACGAGAGCTGCCAGCGCCCTCAAGCAGAGGAGGAGGAGGAGGAG 236

Qy 51 ThrThrAlaIleGluAlaAlaSerSerAspAlaLeuGluAlaLeuAlaAspGlnThrAsp 70

Db 235 TGGGTCAAGAGTCCAGCAGAGCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 176

Qy 71 AlaLeuGlnSerGluGluAlaAlaValValLysAlaAspAsnAlaAlaSerAspAlaLeu 90

Db 175 GCTTTGGAAAGGCAGAGCGGAGGTGTGGCAGCAGAGGAGGAGGAGGAGGAGGAG 116

Qy 91 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaGluValGlnSer 110

Db 115 GAGCAGAGCGGTGTCAGTCAGGAGCTCTGGAGAGGAGGAGGAGGAGGAGGAG 56

Qy 111 AspAsn 112

Db 55 GAGAGC 50

RESULT 29  
A0875428/c  
LOCUS A0875428 494 bp DNA linear GSS 08-NOV-1999  
DEFINITION cerevisiae genomic 5', DNA sequence.

ACCESSION A0875428.1 GI:6287672  
VERSION A0875428.1  
KEYWORDS GSS.  
SOURCE baker's yeast.  
ORGANISM Saccharomyces cerevisiae

REFERENCE  
AUTHORS Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
1 (bases 1 to 494)

TITLE Large-Scale Analysis of the Yeast Genome by Transposon Tagging and  
Gene Disruption  
JOURNAL Unpublished (1999)  
COMMENT Contact: Kumar A

Michael Snyder, Dept. of Mol. Cell. and Dev. Biology  
Yale University  
P.O. Box 208103, New Haven, CT 06520-8103, USA  
Tel: 203 432 9949  
Fax: 203 432 6161  
Email: anuj.kumar@yale.edu

te of mTn-3xHA/lacZ insertion.  
Seq primer: GGCCTCTCTCTTGGCAAGTAC  
Class: Transposon-tagged.  
Location/Qualifiers  
1. 494  
/organism="Saccharomyces cerevisiae"  
/strain="Y2278 - S288C background, cir(0) rho(0)"  
/db\_xref="taxon:4932"  
/clone\_lib="mTn-3xHA/lacZ Insertion Library, strain Y2278"  
/lab\_host="E. coli"  
/note="Vector: pHS56-Sal; A yeast genomic DNA library

without 2 micron or mitochondrial DNA was prepared in  
pHS56-Sal; genomic DNA was size-fractionated (DNA of  
roughly 2-3 kb in length) prior to cloning. This library  
was subsequently mutagenized with a mTn-3xHA/lacZ  
minitransposon containing lacZ, URA3, and tet resistance.

BASE COUNT 133 a 96 c 92 g 173 t  
ORIGIN

## Alignment Scores:

Pred. No.: 6.44 Length: 494  
Score: 78.00 Matches: 19  
Percent Similarity: 56.86% Conservative: 10  
Best Local Similarity: 37.25% Mismatches: 22  
Query Match: 10.04% Indels: 0  
DB: 12 Gaps: 0

US-09-847-539A-6 (1-159) x A0875428 (1-494)

Qy 81 LysAlaAspAsnAlaAlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeu 100

Db 258 AAGCTGAACCTGCCGCCCAAGATGTCCAAACAAGTTGGAGAACCACCAAGATCTTG 199

Qy 101 GlnSerGluGluAlaGluValValGlnSerAspAsnAlaAlaSerAspAlaTrpGluLys 120

Db 198 CAAACAAGGCGCAAGAGTAAAGGAACAAGCTGAAGCTTCTATCGACAACCTAAAAAT 139

Qy 121 AlaAlaThrProIleAlaLeuAspValLysLys 131

Db 138 GAAGCTACTCCAGAAGCTGAACAGGTGAAGAAG 106

RESULT 30  
BI863069/c

LOCUS BI863069 498 bp mRNA linear EST 10-OCT-2001

DEFINITION rm16b06.y1 Meloidogyne arenaria egg pAMP1 v1 Chiapelli McCarter  
Meloidogyne arenaria cDNA 5', mRNA sequence.

ACCESSION BI863069  
VERSION BI863069.1 GI:16004859

KEYWORDS EST.  
SOURCE Meloidogyne arenaria.

ORGANISM Meloidogyne arenaria

REFERENCE Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;  
Tylenchoidea; Heteroderidae; Meloidogyninae; Meloidogyne.

AUTHORS 1 (bases 1 to 498)

McCarter, J., Clifton, S., Chiapelli, B., Pape, D., Martin, J., Wylie, T.,  
Dante, M., Marra, M., Hillier, D., Kucaba, T., Theising, B., Bowers, Y.,  
Gibbons, M., Ritter, E., Bennett, J., Franklin, C., Tsagarelshvili, R.,  
Ronko, I., Kennedy, S., Maguire, L., Beck, C., Underwood, K., Steptoe,  
M., Allen, M., Person, B., Swaller, T., Harvey, N., Schurk, R., Kohn, S.,  
Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and  
Wilson, R.

The Washington Univ. Nematode EST Project, 1999  
Unpublished (1999)

CONTACT: McCarter JP

The Washington Univ. Nematode EST Project, 1999

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@wustl.edu

The library was constructed by Brandi Chiapelli and Dr. James

McCarter (bchiapell@wustl.edu & jmcarter@wustl.edu) at

Washington University, St. Louis. DNA Sequencing by: Washington

University Genome Sequencing Center St. Louis. Nematodes were

provided by Dr. David Bird and Daniel Snyder of North Carolina

State University.

Seq primer: -40RP from Gibco

High quality sequence stop: 415.

Location/Qualifiers

1. 498

/organism="Meloidogyne arenaria"

/db\_xref="taxon:6304"

/clone\_lib="Meloidogyne arenaria egg pAMP1 v1 Chiapelli



```

/db_xref="taxon:7091"
/clone="ws00597"
/clone_lib="Bombyx mori p50(Daizo)"
BASE COUNT      243 a 177 c 220 g 113 t
ORIGIN

Alignment Scores:
Pred. No.:      10.9      Length:      753
Score:          78.00     Matches:    24
Percent Similarity: 44.34%  Conservative: 23
Best Local Similarity: 22.64%  Mismatches: 59
Query Match:    10.04%    Indels:    0
DB:            9         Gaps:      0

US-09-847-539A-6 (1-159) x AU003803 (1-753)
Qy 37 IleAspGluLeuLysGlnAlaIleGluAspLysGluAlaThrThrAlaIleGluAla 56
Db 291 CTGGATCAGCTGCAGCGCCAGCTTAAAGAGTTCAAGTGTGGAACCTCAGCAAGCCACGCTTGACA 350
Qy 57 AlaSerSerAspAlaLeuAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGlu 76
Db 351 GCCCAATGAGATTACAGACTTAAAGAGTTCAAGTGTGGAACCTCAGCAAGCCACGCTTGACA 410
Qy 77 AlaAlaValLysAlaAspAsnAlaAlaSerAspAlaLeuGluAlaLeuAlaAspGln 96
Db 411 AAGCAAGCGATCCAGCCAGCAGTGCAGCGCCGCTCCAGAGAGAGGTAGCGCGGCC-470
Qy 97 ThrAspAlaLeuGlnSerGluGluAlaGluValGlnSerAspAsnAlaAlaSerAsp 116
Db 471 GAGGAGAGCATCAACAGCATCAGCAGGAGCTGGTCCGCCAGTGACAGCGCAAGGAG 530
Qy 117 AlaTrpGluLysAlaAlaThrProIleAlaLeuAspValLysThrLysAspThrLys 136
Db 531 GCCAGGAGCTGGCCCGCAGCAGCCAAAGGAATATGCCGACAAAGCTTCTGAGGCGCCG 590
Qy 137 ProValValLysLysGlu 142
Db 591 CACGAATCAAGAAGAAA 608

RESULT 33
AJ395963
LOCUS          AJ395963      755 bp      mRNA      linear      EST 25-JAN-2001
DEFINITION    AJ395963      dkfz426 Gallus gallus cDNA clone 24m17rl, mRNA sequence.
ACCESSION     AJ395963
VERSION       AJ395963.1 GI:7127417
KEYWORDS      EST.
SOURCE        chicken.
ORGANISM      Gallus gallus
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
               Phasianinae; Gallus.
REFERENCE     1 (bases 1 to 755)
AUTHORS      Adrakhmanov,I., Lodygin,D., Geroth,P., Arakawa,H., Law,A., Plachy
               J., Korn,B. and Buerstedde,J.M.
TITLE        A large database of chickenursal ESTs as a resource for the
               analysis of vertebrate gene function
JOURNAL       Genome Res. 10 (12), 2062-2069 (2000)
MEDLINE       20568495
COMMENT       Contact: Buerstedde JM
               Cellular Immunology
               Heinrich-Pette-Institute
               Martinistr. 52, 20251 Hamburg, Germany
               Email: URL: http://genetics.hpi.uni-hamburg.de/dt40est.html.
               Location/Qualifiers
                 1. 755
                   /organism="Gallus gallus"
                   /strain="CB"
                   /db_xref="taxon:9031"
                   /clone="24m17rl"
                   /clone_lib="dkfz426"
                   /tissue_type="Bursa of Fabricius"
BASE COUNT    315 a 124 c 185 g 131 t

```

---

```

ORIGIN

Alignment Scores:
Pred. No.:      10.9      Length:      755
Score:          78.00     Matches:    22
Percent Similarity: 47.06%  Conservative: 18
Best Local Similarity: 25.88%  Mismatches: 45
Query Match:    10.04%    Indels:    0
DB:            9         Gaps:      0

US-09-847-539A-6 (1-159) x AJ395963 (1-755)
Qy 62 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluAlaAlaValValLys 81
Db 214 GTTGACGATTAGTTCTGGAACCTTGAACAGCTGAACAGCAAGAGCTCATATAAGCAA 273
Qy 82 AlaAspAsnAlaAlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGln 101
Db 274 CAGAGTCAGGCGAGCACCAACAGCAATCGCATCTTAAAGGAGCAAGTCAGTCTTTGGAA 333
Qy 102 SerGluGluAlaGluValValGlnSerAspAsnAlaAlaSerAspAlaTrpGluLysAla 121
Db 334 GCTGAGCAGTGAAGCAACAGGGAATCTTAAAGAACGCGAGAGATGAGCTGTCAGTGAA 393
Qy 122 AlaThrProIleAlaLeuAspValLysThrLysAspThrLysProValValLysLys 141
Db 394 AAGGATTATGGCAGACGAACTAAAGATATTAAAGCCAATCTGCAAGATAGAGAAA 453
Qy 142 GluGluArgGlnAsn 146
Db 454 TACAGAGAGCAAAAT 468

RESULT 34
CNS07CXJ/c
LOCUS          CNS07CXJ      968 bp      DNA      linear      GSS 08-JUL-2001
DEFINITION    T7 end of clone BD0AA007F08 of library BD0AA from strain CBS 94 of
               Candida tropicalis, genomic survey sequence.
ACCESSION     AL439757
VERSION       AL439757.1 GI:12223169
KEYWORDS      GSS.
SOURCE        Candida tropicalis.
ORGANISM      Candida tropicalis
               Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
               Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE     1 (bases 1 to 968)
AUTHORS      Blandin,G., Ozier-Kalogeropoulos,O., Wincker,P., Artiguenave,F. and
               Dujon,B.
TITLE        Genomic exploration of the hemiascomycetous yeasts: 16. Candida
               tropicalis
JOURNAL       FEBS Lett. 487 (1), 91-94 (2000)
MEDLINE       20584726
REFERENCE     2 (bases 1 to 968)
AUTHORS      Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
               Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
               de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
               Malpertuy,A., Neuvéglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
               Saurin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
               Wincker,P. and Weissenbach,J.
TITLE        Genomic exploration of the hemiascomycetous yeasts: 1. A set of
               yeast species for molecular evolution studies
JOURNAL       FEBS Lett. 487 (1), 3-12 (2000)
MEDLINE       20584711
REFERENCE     3 (bases 1 to 968)
AUTHORS      Direct Submission
TITLE        Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,
               2 rue Gaston Cremieux, Cp 5706, 91057 EVRY cedex, FRANCE. (E-mail :
               seqref@genoscope.cns.fr - Web :
               This GSS is part of a random genomic sequencing program of thirteen
               yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
               exiguus, Saccharomyces servazii, Zygosaccharomyces rouxii,
               Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
               lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia

```

angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila, Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.

#### FEATURES

Location/Qualifiers  
 1. .968  
 /organism="Candida tropicalis"  
 /strain="CBS 94"  
 /db\_xref="taxon:5482"  
 /clone="BD0AA007F08"  
 /clone\_lib="BD0AA"  
 /note="end : 77"

BASE COUNT 243 a 231 c 164 g 327 t 3 others

#### ALIGNMENT SCORES:

Pred. No.: 14.9 Length: 968  
 Score: 78.00 Matches: 29  
 Percent Similarity: 48.96% Conservative: 18  
 Best Local Similarity: 30.21% Mismatches: 46  
 Query Match: 10.04% Indels: 3  
 DB: 12 Gaps: 0

US-09-847-539A-6 (1-159) x CNS07CXJ (1-968)

QY 50 AlaThrThrAlaTleGluAlaAlaSerSerAspAlaLeuGluAlaLeuAlaAspGlnThr 69

Db 329 GCACGACAGTATTACCAACGGGAGCAGATCCAGTTCAGTACGATCAGATCAGACGG 270

QY 70 AspAlaLeuGlnSerGluGluAlaAlaValValLysAlaAspAsnAlaAlaSerAspAla 89

Db 269 GATTTCAGTTTCAGATCAATCAGCATCAGTTTCATCTTGACTACGCGTCAGAACCCAGAT 210

QY 90 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluAlaLeuAlaValGln 109

Db 209 TTGGAAGCAGTCGCGTCAGTTTCATCAGTATTAACAGAACAGTCCTGGAATCAGAG 150

QY 110 SerAsp---AsnAlaAlaSerAspAlaTTPGluLysAlaAlaThrProfileAlaLeuAsp 128

Db 149 CCAGATTGTTGTCAGTTCGGATGATCATGTTTCAGAGCACTGCATTAGATTAGAA 90

QY 129 ValLysLysThrLysAspThrLysProValValLysLysGlu 142

Db 89 TCGGAACCAAGAAAGAAATCGGCACCAAGAGTAGAAGCAGAA 48

#### RESULT 35

AQ011873 394 bp DNA linear GSS 25-MAY-2001  
 LMAJFV1\_in07b12.y1 Leishmania major FV1 random genomic library  
 Leishmania major genomic clone LMAJFV1\_in07b12 5' similar to  
 contains element V1-ch1\_type\_II.5 leishmania repetitive element ;,  
 DNA sequence.

#### ACCESSION

VERSION AQ011873.1 GI:6508437

#### KEYWORDS

GSS.

#### SOURCE

Leishmania major.

#### ORGANISM

Eukaryota; Euzlenozoa; Kinetoplastida; Trypanosomatidae;

#### REFERENCE

AUTHORS Akopyants, N.S., Clifton, S.W., Martin, J., Pape, D., Wylie, T., Li, L., Kissinger, J., Roos, D.S., Marra, M., Hillier, L., Chinwalla, A., Blistain, A., Schmitt, A., Person, B., Theising, B., Ritter, E., Ronko, I., Bennett, J., Cole, R., Underwood, K., Cardenas, M., Gibbons, M., Harvey, N., McCann, R., Tsagarisvili, R., Williams, T., Jackson, Y., Bowers, Y., Swaller, T., Waterston, R., Wilson, R. and Beverley, S.M.  
 A survey of the Leishmania major Friedlin strain V1 genome by shotgun sequencing: a resource for DNA microarrays and expression profiling

#### TITLE

Mol. Biochem. Parasitol. 113 (2), 337-340 (2001)

#### JOURNAL

MEDLINE 21192569

#### COMMENT

Contact: Akopyants, NS / Beverley, SM

WashU Leishmania Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: estewartson.wustl.edu

Library construction: Natalia S. Akopyants, Ph.D.  
 Library sequencing by: Washington University Genome Sequencing Center

If using this information please cite:

N.S. Akopyants and S.M. Beverley 'A survey of the Leishmania major Friedlin strain V1 genome by shotgun sequencing' and the Washington University Genome Sequencing Center for information on obtaining clone material please contact: Natalia S. Akopyants Ph.D. (natalia@borcim.wustl.edu) and/or Stephen M. Beverley Ph.D. (beverley@borcim.wustl.edu)

Seq primer: -40RP from Gibco

Class: shotgun

High quality sequence stop: 374.

#### FEATURES

Location/Qualifiers  
 1. .394

/organism="Leishmania major"

/strain="Friedlin strain V1"

/db\_xref="taxon:5664"

/clone="LMAJFV1\_in07b12"

/lab\_host="Leishmania major FV1 random genomic library"

/note="Vector: pZero-2 (Invitrogen)"

Genomic DNA was isolated from stationary phase cells. For

this library, DNA was sheared to give a tight size

distribution of 1-1.5kb fragments, blunt-ended with T4 DNA

polymerase, dephosphorylated with Shrimp Alkaline

Phosphatase and ligated into pZero-2 vector's EcoRV site."

BASE COUNT 50 a 115 c 179 g 50 t

#### ORIGIN

Alignment Scores:  
 Pred. No.: 6.48 Length: 394  
 Score: 77.00 Matches: 22  
 Percent Similarity: 51.95% Conservative: 18  
 Best Local Similarity: 28.57% Mismatches: 37  
 Query Match: 9.91% Indels: 0  
 DB: 12 Gaps: 0

US-09-847-539A-6 (1-159) x AQ011873 (1-394)

QY 46 GluAspLysGluAlaThrThrAlaLeuAlaAlaSerSerAspAlaLeuGluAlaLeu 65

Db 82 GACATGCCCGCGCTGGTTCGCTGGAGGCGAAGCGCGCTCCGAGACCGAGCTG 141

QY 66 AlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLysAlaAspAsnAla 85

Db 142 GCTGCTGCGCGCAGCGATACGCTGCTGAGGAGCGGCTGCGCGAGCGGACGCTGAG 201

QY 86 AlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAla 105

Db 202 CATGCCCGCGCTGCTGCTGCGCTGGAGGCGAAGCGCGTGTGCTCCGAGCGTGGAGCGCT 261

QY 106 GluValValGlnSerAspAsnAlaAlaSerAspAlaTrpGluLysAlaAla 122

Db 262 GCTGCCCTGGAGGCGAAGCGCGCTGCGAGCGGACGCTGCTGCTGCGG 312

#### RESULT 36

AA784607

#### LOCUS

DEFINITION

f2f10a1.r1 Aspergillus nidulans 24hr asexual developmental and

vegetative cDNA lambda zap library Emericella nidulans cDNA clone

f2f10a1 5', mRNA sequence.

ACCESSION AA784607

VERSION AA784607.1 GI:2844775

KEYWORDS EST.

SOURCE Emericella nidulans.

ORGANISM Emericella nidulans.

Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;

AA784607 425 bp mRNA linear EST 29-JUL-1998

f2f10a1.r1 Aspergillus nidulans 24hr asexual developmental and

vegetative cDNA lambda zap library Emericella nidulans cDNA clone

f2f10a1 5', mRNA sequence.

ACCESSION AA784607

VERSION AA784607.1 GI:2844775

KEYWORDS EST.

SOURCE Emericella nidulans.

ORGANISM Emericella nidulans.

Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;



US Department of Agriculture, Agriculture Research Service, Pacific  
West Area, Western Regional Research Center  
800 Buchanan Street, Albany, CA 94710, USA  
Tel: 5105595773  
Fax: 5105595818  
Email: candersnpw.usda.gov  
Sequence have been trimmed to remove vector sequence and low  
quality sequence with phred score less than 20  
Seq primer: Stratagene SK primer.

**FEATURES**  
**SOURCE**

```

/organism="Triticum aestivum"
/cultivar="Chinese Spring"
/db_xref="taxon:4365"
/clone="WHE2336_F03_L06"
/clone_lib="Wheat pre-anthesis spike cDNA library"
/tissue_type="Spike before anthesis"
/dev_stage="Adult plant"
/lab_host="E. coli SOLR"
/notes=Vector: Lambda Uni-ZAP XR, excised phagemid
Site_1: EcoRI; Site_2: XhoI; Plants were grown in
greenhouse. Whole spike with awns trimmed, white
and yellow anther were collected and total RNA, an
poly(A) RNA were prepared, a cDNA library was made
the cDNA clones were in vivo excised to give pBLu
phagemids in the TJ Close lab (Choi, Close, Fent
the University of California, Riverside. Plasmid
preparations and DNA sequencing were performed in
Anderson lab (all other authors)".

BASE COUNT      135 a      262 c      188 g      78 t
ORIGIN

Alignment Scores:
Pred. No.:      12.4      Length:      663
Score:          77.00      Matches:     32
Percent Similarity: 39.32%      Conservative: 14
Best Local Similarity: 27.35%      Mismatches:  70
Query Match:     9.91%      Indels:      1
DB:              10      Gaps:        0

```

RESULT 39	675 bp	mRNA	linear	EST 19-JAN-1999
AU003869	Bombyx mori p50(Daizo)	Bombyx mori	cDNA clone	ws00687,
LOCUS	mRNA sequence.			
DEFINITION				
ACCESSION	AU003869			
VERSION	AU003869.1			
	GI:4161240			

**KEYWORDS**  
**SOURCE**  
**ORGANISM**

EST.  
domestic silkworm.

Bombyx mori  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia  
; Bombycoidea; Bombycidae; Bombyx.

REFERENCE  
1 (bases 1 to 675)  
AUTHORS Mita, K., Morimyo, M., Shimada, T., Okano, K. and Maeda, S.  
TITLE Establishment of cDNA database of Bombyx mori  
JOURNAL Unpublished (1999)  
COMMENT Contact: Mita, K

FEATURE SOURCE

```

FEATURES
source
1. 675
/organism="Bombyx mori"
/strain="p50(Daizo)"
/db_xref="taxon:7091"
/clone="ws00667"
211 a 164 c 193 a 1
BASE COUNT

```

Alignment Scores:	
Pred. No.:	12.7
Score:	77.00
Percent Similarity:	44.34%
Best Local Similarity:	22.64%
Query Match:	9.91%
DB:	9
Length:	675
Matches:	24
Conservative:	23
Mismatches:	59
Indels:	0
Gaps:	0

RESULT 40					
AZ948693					
LOCUS	AZ948693	723 bp	DNA	linear	GSS 27-APR-2001
DEFINITION	2M0211018R Mouse 10kb plasmid UUGC2M library Mus musculus genomic clone UUGC2M0211018 R, DNA sequence.				

ACCESSION	AZ948693
VERSION	AZ948693.1
KEYWORDS	GI:13819920 GSS.

SOURCE house mouse.  
ORGANISM Mus musculus.  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 723)

AUTHORS

Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,  
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly  
M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.  
and Wright,D., Weiss,R.

TITLE

Mouse whole genome scaffolding with paired end reads from 10kb

JOURNAL  
COMMENT

plasmid inserts  
Unpublished (2000)  
Contact: Robert B. Weiss  
University of Utah Genome Center  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0211 row: 0 column: 18  
Seq primer: CACACAGGAACAGCTATGACC  
Class: plasmid ends  
High quality sequence stop: 723.

FEATURES

source

1..723  
Location/Qualifiers  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGC2M0211018"  
/clone\_lib="Mouse 10kb plasmid UUGC2M library"  
/sex="Female"  
/lab\_host="E. coli strain XL10-Gold, Tl-resistant, F-"  
/note="Vector: PWD42nv; Purified genomic DNA from M.  
musculus C57BL/6J (female) was obtained from the Jackson  
Laboratory Mouse DNA Resource  
(http://www.jax.org/resources/documents/dnares/). The DNA  
was hydrodynamically sheared by repeated passage through a  
0.005 inch orifice at constant velocity. The sheared DNA  
was blunt end-repaired with T4 DNA polymerase and T4  
polynucleotide kinase. Adaptor oligonucleotides were  
ligated to the blunt ends in high molar excess. The  
adaptored DNA was purified and size-selected for a 9.5 to  
10.5 kb range using preparative agarose gel  
electrophoresis. Vector DNA was prepared from a derivative  
of pWD42 (g1473211419b1AF129072.1), a copy-number  
inducible derivative of plasmid R1. The vector was ligated  
with adaptors complementary to the insert adaptors and  
purified. The sheared, adaptored mouse DNA was annealed to  
adaptored vector DNA, and transformed into  
chemically-competent E. coli XL10-Gold (Stratagene) cells  
and selected for ampicillin resistance."

BASE COUNT 245 a 133 c 254 g 91 t

ORIGIN

Alignment Scores:

Pred. No.:	13.8	Length:	723
Score:	77.00	Matches:	25
Percent Similarity:	46.15%	Conservative:	17
Best Local Similarity:	27.47%	Mismatches:	49
Query Match:	9.91%	Indels:	0
DB:	12	Gaps:	0

US-09-847-539A-6 (1-159) x AZ948693 (1-723)

QY	31	ArgAsnGluGluArgAlaIleAspGluLeuLysLysGlnAlaIleGluAspLysGluAla	50
Db	6	CGCGAGCGCAGAGGAGCGAGGCGAGGCGAGGCGAGGCGAGGCGAGGCGAGGCGCA	65
QY	51	ThrThrAlaIleGluAlaSerSerAspAlaLeuGluAlaLeuAlaAspGlnThrAsp	70
Db	66	GAGGAGGCGAGGAGGCGAGGAGGCGAGGCGAGGCGAGGCGAGGCGAGGAGGAG	125
QY	71	AlaLeuGlnSerGluAlaAlaValValLysAlaAspAsnAlaAlaSerAspAlaLeu	90
Db	126	GCAGAGGAGGCGAGGAGGCGAGGAGGCGAGGAGGCGAGGAGGCGAGGAGGCGAG	185

QY	91	GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaGluValValGlnSer	110
Db	186	GCAGAGGCGCAGGCGAGGCGAGGCGAGGCGAGGCGAGGCGAGGCGAGGCGAGGCA	245
QY	111	AspAsnAlaAlaSerAspAlaTrpGluLysAla	121
Db	246	GCAGAGGCGCAGGCGAGGCGAGGCGAGGCGAGGCGAGGCGAGGCGAGGCA	278

Search completed: October 13, 2002, 03:33:06  
Job time : 1894.74 secs

THIS PAGE BLANK (USPTO)



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

# OM protein - protein search, using sw model

Run on: October 13, 2002, 01:12:51 ; Search time 22.9572 Seconds  
(without alignments)  
268.169 Million cell updates/sec

Title: US-09-847-539A-6  
Perfect score: 777  
Sequence: 1 VDSPISQPRIIPNGGTLTNL.....KKEERQNVNTLPTTGESNP 159

Scoring table: BLOSUM62  
Gapop 60.0 , Gapext 60.0

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	155	19.9	448	1 SPGL_STRSP	P06654 streptococ
2	155	19.9	593	1 SPGL_STRSP	P19909 streptococ
3	99	12.7	774	1 STFL_LAMBD	P03764 bacterioph
4	88	11.3	407	1 IE68_HSVSA	Q01042 herpesviru
5	85	10.9	1120	1 STFR_ECOLI	P76072 escherichia
6	80	10.3	269	1 TIR3_YEAST	P40552 saccharomyc
7	80	10.3	347	1 TOLA_PSEAE	P50600 pseudomonas
8	79	10.2	389	1 YM82_YEAST	Q04951 saccharomyc
9	78	10.2	516	1 P54_ENTFC	P13692 enterococcu
10	78	10.0	771	1 CALD_CHICK	P12957 gallus gall
11	76	9.8	490	1 CE05_ECOLI	Q47500 escherichia
12	76	9.8	490	1 CE10_ECOLI	Q47125 escherichia
13	76	9.8	1290	1 XPCX_XENLA	P50532 xenopus lae
14	75	9.7	406	1 SR40_YEAST	P32583 saccharomyc
15	75	9.7	3210	1 CENF_HUMAN	P49454 homo sapien
16	73	9.4	1189	1 SCIL_CHICK	Q90988 gallus gall
17	73	9.4	1278	1 DINA_HUMAN	Q44203 homo sapien
18	73	9.4	1396	1 VLTF_BPT5	P13390 bacterioph
19	72	9.3	464	1 AK15_RAT	P24587 rattus norv
20	72	9.3	488	1 CVL2_BOVIN	Q28092 bos taurus
21	72	9.3	2230	1 GOG4_HUMAN	Q13439 homo sapien
22	71	9.1	217	1 GPPE_HALN1	Q9hrv0 halobacteri
23	71	9.1	439	1 DDY7_HUMAN	Q15320 homo sapien
24	71	9.1	1238	1 SECC_RHOCA	Q68032 rhodobacter
25	70	9.0	211	1 T2D9_HUMAN	Q15544 homo sapien
26	70	9.0	211	1 TUB8_SOLTU	P33191 solanum tub
27	70	9.0	421	1 TOLA_ECOLI	P19934 escherichia
28	70	9.0	433	1 HTR2_HALVA	P42258 haloscula
29	70	9.0	857	1 NPM_CHICK	P16053 gallus gall
30	70	9.0	1391	1 MST1_DROHY	Q88696 drosophil
31	69	8.9	198	1 HBHA_MYCTU	Q11142 mycobacteri
32	69	8.9	500	1 GAR2_SCHPO	P41891 schizosacch
33	69	8.9	889	1 HS97_STRPU	Q06068 strongyloce

34	69	8.9	1131	1 AC15_MOUSE	P35601 mus musculu
35	69	8.9	1281	1 DYNA_MOUSE	Q08788 mus musculu
36	69	8.9	1423	1 FRUA_STRMU	Q03174 streptococ
37	69	8.9	1528	1 SPAA_STRDO	P21979 streptococ
38	68	8.8	232	1 YJFJ_ECOLI	P39292 escherichia
39	68	8.8	643	1 DP3X_ECOLI	P06710 escherichia
40	67	8.6	654	1 DNAK_CHLMU	P56836 chlamydia m
41	67	8.6	756	1 Y328_MYCGE	Q04919 mycoplasma
42	67	8.6	763	1 HTR2_HALN1	Q9hp81 halobacteri
43	67	8.6	764	1 HTR2_HALSA	P71410 halobacteri
44	67	8.6	1399	1 RPOC_PSEAE	Q9hwc9 pseudomonas
45	67	8.6	1407	1 RPOC_BUCAI	P57145 buchnera ap

## ALIGNMENTS

RESULT 1  
SPGL\_STRSP STANDARD; PRT; 448 AA.  
AC P06654:  
DT 01-JAN-1988 (Rel. 06, Created)  
DT 01-JAN-1988 (Rel. 06, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Immunoglobulin G binding protein G precursor (IGG binding protein G).  
GN SPC.  
OS Streptococcus sp. (Lancefield group G).  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1306;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=863041178; PubMed=3745123;  
RA Fahnestock S.R., Alexander P., Nagle J., Filpula D.;  
RT "Gene for an immunoglobulin-binding protein from a group G streptococcus.";  
RL J. Bacteriol. 167:870-880(1986).  
RN [2]  
RP X-RAY CRYSTALLOGRAPHY (2.07 ANGSTROMS) OF 228-282.  
RX MEDLINE=94213848; PubMed=8161530;  
RA Gallagher T., Alexander P., Bryan P., Gilliland G.L.;  
RT "Two crystal structures of the B1 immunoglobulin-binding domain of streptococcal protein G and comparison with NMR.";  
RN [3]  
RX Biochemistry 33:4721-4729(1994).  
RP X-RAY CRYSTALLOGRAPHY (1.1 ANGSTROMS) OF 293-351.  
RX MEDLINE=95055731; PubMed=7966308;  
RA Derrick J.P., Wigley D.B.;  
RT "The third IgG-binding domain from streptococcal protein G. An analysis by X-ray crystallography of the structure alone and in a complex with Fab.";  
RL J. Mol. Biol. 243:906-918(1994).  
RN [4]  
RP X-RAY CRYSTALLOGRAPHY (3.5 ANGSTROMS) OF 297-352.  
RX MEDLINE=95308043; PubMed=7788293;  
RA Sauer-Eriksson A.E., Kleywegt G.J., Uhlen M., Jones T.A.;  
RT "Crystal structure of the C2 fragment of streptococcal protein G in complex with the Fc domain of human IgG.";  
RL Structure 3:265-278(1995).  
RN [5]  
RP X-RAY CRYSTALLOGRAPHY (1.1 ANGSTROMS) OF 292-352.  
RA Butterworth S., Lanzin V.S., Wigley D.B., Derrick J.P., Wilson K.S.;  
RT Submitted (APR-1997) to the PDB data bank.  
RN [6]  
RP STRUCTURE BY NMR OF 298-351.  
RX MEDLINE=91335209; PubMed=1871600;  
RA Gronenborn A.M., Filpula D.R., Essig N.Z., Achari A., Whitlow M., Wingfield P.T., Clore G.M.;  
RT "A novel, highly stable fold of the immunoglobulin binding domain of streptococcal protein G.";  
RL Science 253:657-661(1991).  
CC -!- FUNCTION: BINDS TO THE CONSTANT FC REGION OF IGG WITH HIGH AFFINITY.

```
CC CC -!- SUBCELLULAR LOCATION: Type I membrane protein. Cell wall.
CC CC -!- SIMILARITY: TO OTHER STREPTOCOCCAL AND STAPHYLOCOCCAL PROTEINS
CC CC IN THE REGION OF THE MEMBRANE ANCHOR.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M13825; AAA03664.1; -
CC PIR: A24496; A24496.
CC PDB: 1PGA; 30-APR-94.
CC PDB: 1PGX; 15-JUL-92.
CC PDB: 1IGC; 03-JUN-95.
CC PDB: 1IGD; 01-NOV-94.
CC PDB: 2IGD; 29-JUL-98.
CC PDB: 1FGC; 20-JUL-95.
CC PDB: 1GB1; 15-APR-93.
CC InterPro: IPR002988; GA.
CC InterPro: IPR001899; Gram_pos_anchor.
CC InterPro: IPR000724; IgG_bind_B.
CC Pfam: PF01468; GA; 2.
CC Pfam: PF00746; Gram_pos_anchor; 1.
CC Pfam: PF01378; IgG_binding_B; 2.
CC PRINTS: PR00015; GPOSANCHOR.
CC PROSITE: PS00343; GRAM_POS_ANCHORING; 1.
CC IgG-binding protein; Repeat; Transmembrane; Cell wall; Signal;
CC 3D-structure.
CC SIGNAL 1 33 IMMUNOGLOBULIN G BINDING PROTEIN G.
CC CHAIN 34 448 EXTRACELLULAR.
CC DOMAIN 423 443
CC TRANSMEM 444 448
CC DOMAIN 69 111 CYTOPLASMIC.
CC DOMAIN 104 215 ALA-RICH.
CC REPEAT 104 140 2 X 37 AA REPEATS.
CC REPEAT 179 215 1-1.
CC DOMAIN 228 352 1-2.
CC REPEAT 228 352 2 X 55 AA REPEATS.
CC REPEAT 238 352 2-1.
CC REPEAT 238 352 2-2.
CC DOMAIN 386 410 5 X 5 AA REPEATS OF [DE]-D-A-K-K.
CC DOMAIN 414 419 CONSERVED IN GRAM-POSITIVE COCCI SURFACE
CC PROTEINS.
CC SEQUENCE 448 AA; 47567 MW; A0759060C8F956CA CRC64;
Query Match 19.9%; Score 155; DB 1; Length 448;
Best Local Similarity 56.1%; Pred. No. 1.3e-09;
Matches 37; Conservative 4; Mismatches 25; Indels 0; Gaps 0;
QY 1 VDSPIEQPIIPNGGTLNLIGNAPEKALRNERRAIDELKKQAIEDKKAATTAIEAASD 60
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 34 VDSPIEDTPIIRNGBELNLGNSETTLALRNESATADLTAAAVADTVAAAAAENAGAA 93
QY 61 ALEALA 66
|||
Db 94 AWEAAA 99
RESULT 2
SPG2_STRSP STANDARD; PRT; 593 AA.
AC SPG2_STRSP
AC P19909;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Immunoglobulin G binding protein G precursor (IGG binding protein G).
GN SPG.
OS Streptococcus sp. (strain G148).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
```

```
OC Streptococcus.
OX NCBI_TaxID=1306;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-G148;
RX MEDLINE=88029445; PubMed=3665928;
RA Olsson A., Eliasson M., Guss B., Nilsson B., Hellman U.,
Lindberg M., Uhlen M.;
RT "Structure and evolution of the repetitive gene encoding
streptococcal protein G.";
RL Eur. J. Biochem. 168:319-324(1987).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-CX7805;
RX MEDLINE=88015586; PubMed=3658689;
RA Filipula D., Alexander P., Fahnestock S.R.;
RT "Nucleotide sequence of the protein G gene from Streptococcus GX7805,
and comparison to previously reported sequences.";
RL Nucleic Acids Res. 15:7210-7210(1987).
RN [3]
RP SEQUENCE OF 114-593 FROM N.A.
RC STRAIN-G148;
RX MEDLINE=86300657; PubMed=3017704;
RA Guss B., Eliasson M., Olsson A., Uhlen M., Frej A.-K., Joernvall H.,
Flock J.-I., Lindberg M.;
RT "Structure of the IgG-binding regions of streptococcal protein G.";
RL EMBO J. 5:1567-1575(1986).
RN [4]
RP STRUCTURE BY NMR OF 371-427.
RC STRAIN-G148;
RX MEDLINE=98290449; PubMed=9628485;
RA Malakauskas S.M., Mayo S.L.;
RT "Design, structure and stability of a hyperthermophilic protein
variant.";
RL Nat. Struct. Biol. 5:470-475(1998).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein. Cell wall.
CC -!- SIMILARITY: TO OTHER STREPTOCOCCAL AND STAPHYLOCOCCAL PROTEINS
CC IN THE REGION OF THE MEMBRANE ANCHOR.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X06173; CAA29540.1; -
CC EMBL: Y00428; CAA68489.1; -
CC PIR: S00128; S00128.
CC PIR: A26314; A26314.
CC PDB: 1GB4; 22-JUL-98.
CC InterPro: IPR002988; GA.
CC InterPro: IPR001899; Gram_pos_anchor.
CC InterPro: IPR000724; IgG_bind_B.
CC Pfam: PF01468; GA; 3.
CC Pfam: PF00746; Gram_pos_anchor; 1.
CC Pfam: PF01378; IgG_binding_B; 3.
CC PRINTS: PR00015; GPOSANCHOR.
CC PROSITE: PS00343; GRAM_POS_ANCHORING; 1.
CC IgG-binding protein; Repeat; Transmembrane; Cell wall; Signal;
CC 3D-structure.
CC SIGNAL 1 33 IMMUNOGLOBULIN G BINDING PROTEIN G.
CC CHAIN 34 593 EXTRACELLULAR (POTENTIAL).
CC DOMAIN 34 567 MEMBRANE ANCHOR.
CC TRANSMEM 568 588 CYTOPLASMIC (POTENTIAL).
CC DOMAIN 589 593 ALA-RICH.
CC DOMAIN 59 111 3 X 37 AA REPEATS.
CC REPEAT 104 290 1-1.
CC REPEAT 104 290 1-2.
CC REPEAT 179 215 1-3.
CC REPEAT 254 290
```

```
FT DOMAIN 303 427 2 X 55 AA REPEATS.
FT REPEAT 303 357 2-1.
FT REPEAT 373 427 2-2.
FT DOMAIN 531 555 5 X 5 AA REPEATS OF [DE]-D-A-K-K.
FT DOMAIN 559 564 CONSERVED IN GRAM-POSITIVE COCCI SURFACE
FT PROTEINS.
SQ SEQUENCE 593 AA; 63291 MW; 048BAA760D5B2920 CRC64;

Query Match 19.9%; Score 155; DB 1; Length 593;
Best Local Similarity 56.18; Pred. No. 1.7e-09;
Matches 37; Conservative 4; Mismatches 25; Indels 0; Gaps 0;

QY 1 VDSPIEPIIPNGGTLNLGNAPKALNEERAIDELKQAIEDKEATTAIEAASSD 60
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 34 VDSPIEDPIIRNGGELTNLGNSETTLALRNEESATADLTAAAVADTVAAAAENAGAA 93

QY 61 ALEALA 66
| | | |
Db 94 AWEAAA 99

RESULT 3
STF_LAMB
ID STF_LAMB STANDARD; PRT; 774 AA.
AC P03764; P03745;
DT 21-JUL-1986 (Rel. 01, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Side tail fiber protein.
GN STF.
OS Bacteriophage lambda.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae;
OC Lambda phage group.
OX NCBI_TaxID=10710;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=83189071; PubMed=6221115;
RA Sanger F., Coulson A.R., Hong G.F., Hill D.F., Petersen G.B.;
RT "Nucleotide sequence of bacteriophage lambda DNA.";
RL J. Mol. Biol. 162:729-773(1982).
RN [2]
RP IDENTIFICATION AS STF.
RX MEDLINE=92165720; PubMed=1531648;
RA Haggard-Ljungquist E., Halling C., Calendar R.;
RT "DNA sequences of the tail fiber genes of bacteriophage P2: evidence
for horizontal transfer of tail fiber genes among unrelated
bacteriophages.";
RL J. Bacteriol. 174:1462-1477(1992).
RN [3]
RP RECONSTRUCTION OF STF.
RX MEDLINE=93068310; PubMed=1439823;
RA Hendrix R.W., Duda R.L.;
RT "Bacteriophage lambda P2: not the mother of all lambda phages.";
RL Science 258:1145-1148(1992).
CC -1- MISCELLANEOUS: The common laboratory strain of bacteriophage
lambda; lambda P2; carries a frameshift mutation relative to Ur-
lambda, the original isolate. The Ur-lambda virions have thin,
jointed tail fibers (side tail fibers) that are absent from lambda
wild type. Relative to lambda P2, Ur-lambda has expanded
receptor specificity and adsorbs to E. coli cells more rapidly.
CC -1- SIMILARITY: BELONGS TO THE TAIL FIBER FAMILY.
CC -1- CAUTION: THIS IS A CONCEPTUAL TRANSLATION; A FRAMESHIFT WAS
CORRECTED IN POSITION TO 396 TO RECREATE THE ORIGINAL STF PROTEIN.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
```

```
DR EMBL; J02459; AAA96557.1; ALT_FRAME.
DR PIR; A04389; QXBP1L.
DR PIR; A04370; QXBP2L.
DR InterPro; IPR004089; Chemotaxis_transducer.
KW Fiber protein.
SQ SEQUENCE 774 AA; 77527 MW; CDD1DF85E919123B CRC64;

Query Match 12.7%; Score 99; DB 1; Length 774;
Best Local Similarity 32.98; Pred. No. 0.0033;
Matches 26; Conservative 16; Mismatches 37; Indels 0; Gaps 0;

QY 50 ATTAIEAASSDALEALADQTDALQSEEAAYVKADNAASDALEADQTDALQSEEAAYVQ 109
| : : | | : : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 190 AESKNAATASAGAAKTSETNAASQQAATSAATKASEAATSDAVASKEAAKSS 249

QY 110 SDNAASDAWEKAATPIALD 128
| | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 250 ETNASSAGRAASSATAAE 268

RESULT 4
IE68_HVSA
ID IE68_HVSA STANDARD; PRT; 407 AA.
AC Q01042;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-APR-1993 (Rel. 25, Last annotation update)
DE Immediate-early protein.
GN 73 OR ECUF1.
OS Herpesvirus saimiri (strain 11).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Rhadinovirus.
OX NCBI_TaxID=10383;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92333686; PubMed=1321287;
RA Albrecht J.-C., Nicholas J., Biller D., Cameron K.R., Biesinger B.,
RA Newman C., Wittmann S., Craxton M.A., Coleman H., Fleckenstein B.,
RA Honess R.W.;
RT "Primary structure of the herpesvirus saimiri genome.";
RL J. Virol. 66:5047-5058(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92230228; PubMed=1314457;
RA Nicholas J., Cameron K.R., Coleman H., Newman C., Honess R.W.;
RT "Analysis of nucleotide sequence of the rightmost 43 kbp of
herpesvirus saimiri (HVS) L-DNA: general conservation of genetic
organization between HVS and Epstein-Barr virus.";
RL Virology 188:296-310(1992).
CC -1- SIMILARITY: BELONGS TO A FAMILY THAT GROUP TOGETHER HSV-1 AND
HSV-2 IE-68 (US1), EBV-1 65, EBV-4 (ORF4), PRV RSP40, AND VZV 63.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
```

```
DR EMBL; X64346; CAA45696.1; -.
DR EMBL; M86409; AAA46149.1; -.
DR EMBL; S76368; AAB21116.1; -.
DR PIR; G36813; EDHEQ3.
DR HSP; S20244; S20244.
DR HSP; P53041; 1A17.
KW Early protein.
FT DOMAIN 60 241 GLU-RICH (ACIDIC).
SQ SEQUENCE 407 AA; 46617 MW; FFD399CA82CE136C CRC64;

Query Match 11.3%; Score 88; DB 1; Length 407;
Best Local Similarity 20.04; Pred. No. 0.028;
Matches 25; Conservative 31; Mismatches 69; Indels 0; Gaps 0;
```









Matches	22;	Conservative	15;	Mismatches	32;	Indels	0;	Gaps	0;
QY	39	ELKKQALIEDKATTAIEAASDALEALADQTDALQSEAAVYVADNAASDALEALADQTD	98						
Db	163	ERQRDEIARQQAETAHLMAEAEEAEKKNQDSDLDEHRAVEAEKKAELAKAEAS	222						
QY	99	ALQSEAEAV	107						
Db	223	DVQSKQAIV	231						
RESULT 13									
XCPC_XENLA									
ID	XCPC_XENLA	STANDARD;	PRT;	1290	AA.				
AC	P50532;								
DT	01-OCT-1996	(Rel. 34, Created)							
DT	01-OCT-1996	(Rel. 34, Last sequence update)							
DT	15-JUL-1998	(Rel. 36, Last annotation update)							
DE	Chromosome assembly protein XCAP-C.								
GN	XCAP-C.								
OS	Xenopus laevis (African clawed frog).								
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;								
OC	Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;								
OC	Xenopodinae; Xenopus.								
OX	NCBI_TaxID=8355;								
[1]									
RP	SEQUENCE FROM N.A.								
RP	MEDLINE=95042742; PubMed=7954811;								
RA	Hirano T., Mitchison T.J.								
RX	"A heterodimeric coiled-coil protein required for mitotic chromosome								
RT	condensation in vitro."								
RL	Cell 79:449-458(1994).								
CC	-1- FUNCTION: REQUIRED FOR BOTH ASSEMBLY AND STRUCTURAL MAINTENANCE OF								
CC	MITOTIC CHROMOSOMES.								
CC	-1- SUBUNIT: ASSOCIATES WITH XCAP-E PROBABLY AS HETERODIMER.								
CC	-1- SUBCELLULAR LOCATION: NUCLEAR. DURING CHROMOSOME ASSEMBLY IN								
CC	MITOTIC EXTRACTS, XCAP-C/E WAS RECRUITED TO THE CHROMATIN AND								
CC	FORMED A DISCRETE INTERNAL STRUCTURE WITHIN ASSEMBLED CHROMOSOMES.								
CC	-1- DOMAIN: CONSISTS OF TWO PUTATIVE CENTRAL COILED-COIL REGIONS								
CC	FLANKED BY PUTATIVE GLOBULAR REGIONS AT THE N- AND C-TERMINUS.								
CC	-1- SIMILARITY: BELONGS TO THE SMC FAMILY.								
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration								
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -								
CC	the European Bioinformatics Institute. There are no restrictions on its								
CC	use by non-profit institutions as long as its content is in no way								
CC	modified and this statement is not removed. Usage by and for commercial								
CC	entities requires a license agreement (see <a href="http://www.isb-sib.ch/announcement/">http://www.isb-sib.ch/announcement/</a>								
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).								
CC	-----								
CC	EMBL: U13673; AAA64679.1; --								
DR	InterPro: IPR003439; ABC_transportr.								
DR	InterPro: IPR003405; SMC_C.								
DR	InterPro: IPR003395; SMC_N.								
DR	Pfam: PF02483; SMC_C; 1.								
DR	Pfam: PF02463; SMC_N; 1.								
KW	Mitosis; ATP-binding; Coiled coil; Nuclear protein.								
FT	NE_BIND 107 114								
FT	DOMAIN 264 594								
FT	DOMAIN 764 1027								
FT	DOMAIN 1094 1129								
FT	DOMAIN 1263 1290								
FT	DOMAIN 60 65								
FT	DOMAIN 747 750								
FT	DOMAIN 841 844								
FT	DOMAIN 1196 1220								
FT	DOMAIN 1290 AA; 14698 MW; 2931249924FE90F6 CRC64;								
SEQ	SEQUENCE								
Query Match	9.8%;	Score 76;	DB 1;	Length 1290;					
Best Local Similarity	21.3%;	Pred. No. 2;							
Matches 16;	Conservative 24;	Mismatches 35;	Indels 0;	Caps 0;					







CC TRANSPORT OF VESICLES AND ORGANELLES.  
CC P150 (GLUED) BINDS DIRECTLY TO MICROTUBULES AND TO CYTOPLASMIC  
CC DYNEIN.  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; P150 (SHOWN HERE) AND P135;  
CC ARE PRODUCED BY ALTERNATIVE SPLICING.  
CC -!- TISSUE SPECIFICITY: BRAIN.  
CC -!- PTM: PHOSPHORYLATED.  
CC -!- SIMILARITY: BELONGS THE DYNACTIN 150 KDA SUBUNIT FAMILY.  
CC -!- SIMILARITY: CONTAINS 1 CAP-GLY DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AF064205; AAD55811.1; .  
DR EMBL; AF064203; AAD55811.1; JOINED.  
DR EMBL; AF064204; AAD55811.1; JOINED.  
DR EMBL; X98801; CAA67333.1; .  
DR EMBL; AF066947; AAD03694.1; .  
DR EMBL; AF086927; AAD03694.1; JOINED.  
DR EMBL; AF086928; AAD03694.1; JOINED.  
DR EMBL; AF086929; AAD03694.1; JOINED.  
DR EMBL; AF086930; AAD03694.1; JOINED.  
DR EMBL; AF086931; AAD03694.1; JOINED.  
DR EMBL; AF086932; AAD03694.1; JOINED.  
DR EMBL; AF086933; AAD03694.1; JOINED.  
DR EMBL; AF086934; AAD03694.1; JOINED.  
DR EMBL; AF086935; AAD03694.1; JOINED.  
DR EMBL; AF086936; AAD03694.1; JOINED.  
DR EMBL; AF086937; AAD03694.1; JOINED.  
DR EMBL; AF086938; AAD03694.1; JOINED.  
DR EMBL; AF086939; AAD03694.1; JOINED.  
DR EMBL; AF086940; AAD03694.1; JOINED.  
DR EMBL; AF086941; AAD03694.1; JOINED.  
DR EMBL; AF086942; AAD03694.1; JOINED.  
DR EMBL; AF086943; AAD03694.1; JOINED.  
DR EMBL; AF086944; AAD03694.1; JOINED.  
DR EMBL; AF086945; AAD03694.1; JOINED.  
DR EMBL; AF086946; AAD03694.1; JOINED.  
DR MIM; 601143; .  
DR InterPro; IPR000938; CAP-Gly.  
DR Pfam; PF01302; CAP\_GLY; 1.  
DR PROSITE; PS00845; CAP\_GLY\_1; 1.  
DR PROSITE; PS0245; CAP\_GLY\_2; 1.  
KW Motor protein; Microtubules; Dynein; Coiled coil; Cytoskeleton;  
KW Alternative splicing; Phosphorylation.  
FT DOMAIN 48 90  
FT CAP-Gly.  
FT SER-RICH.  
FT COILED COIL (POTENTIAL).  
FT COILED COIL (POTENTIAL).  
FT COILED COIL (POTENTIAL).  
FT COILED COIL (POTENTIAL).  
FT MAKRHYVSRTPSSRMSAEASAPLRVGSRVVEVIGKGR  
FT GVAIVGATFLFATGKVGVLDEAKGNDGTGGRKRYFTCD  
FT EHGIFVRSQIQVFDGADFTSPETPDSSASKVLKREGTD  
FT TTKATSKL -> MMRQ (IN ISOFORM P135).  
FT A -> P.  
FT /FTID-VAR\_001373.  
FT S -> N (IN REF. 2 AND 3).  
FT MISSING (IN REF. 2 AND 3).  
FT D -> V (IN REF. 2 AND 3).  
FT VARSPLIC 1 131  
FT  
FT VARIANT 163 163  
FT  
FT CONFLICT 10 10  
FT CONFLICT 132 138  
FT CONFLICT 712 712  
FT CONFLICT 712 712  
FT SEQUENCE 1278 AA; 141694 MW; 6DCEA5E67856E4BC CRC64;  
Query Match 9.4%; Score 73; DB 1; Length 1278;  
Best Local Similarity 24.4%; Pred. No. 4.1;  
Matches 19; Conservative 20; Mismatches 39; Indels 0; Gaps 0;  
QY 38 DELKKQAIEDKETTATIEAASSDALEALADQTDALQSEAAVVKADNAASDALEALADQTD 97

Db 964 EELSEANVRLSLEKLDAAKADADERIEKVQTRLEETQALLRKKEKEFEETMDALQADI 1023  
QY 98 DALQSEEAERVQSDNAAS 115  
Db 1024 DQLEAKAEKORLNSQS 1041  
RESULT 18  
VLTF\_BPT5  
ID VLTF\_BPT5 STANDARD; PRT; 1396 AA.  
AC P13390; O48502;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE L-shaped tail fiber protein (LTF protein).  
GN LTF.  
OS Bacteriophage T5.  
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.  
OX NCBI\_TaxID=10726;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=95309401; PubMed=7789514;  
RA Kaliman A.V., Kulshin V.E., Shiyapnikov M.G., Ksenzenko V.N.,  
RA Kryukov V.M.;  
RT "The nucleotide sequence of the bacteriophage T5 ltf gene.";  
RL FEBS Lett. 366:46-48(1995).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Kaliman A.V.;  
RL Submitted (JAN-1998) to the EMBL/Genbank/DDBJ databases.  
RN [3]  
RP PRELIMINARY PARTIAL SEQUENCE FROM N.A.  
RX MEDLINE=88289370; PubMed=3267228;  
RA Kaliman A.V., Kryukov V.M., Bayev A.A.;  
RT "The nucleotide sequence of bacteriophage T5 DNA at the region  
RT between early and late genes.";  
RL Nucleic Acids Res. 16:6230-6230(1988).  
CC -!- FUNCTION: NONESSENTIAL PROTEIN THAT MEDIATES BINDING TO THE  
CC POLYMANNOSE O ANTIGEN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X69460; CAA49220.1; .  
DR EMBL; AJ001191; CAA04591.1; .  
DR PIR; S01982; S01982.  
KW Late protein.  
FT CONFLICT 986 986 V -> A (IN REF. 2).  
FT SEQUENCE 1396 AA; 147989 MW; 18CD2192F65FFFC1 CRC64;  
Query Match 9.4%; Score 73; DB 1; Length 1396;  
Best Local Similarity 29.5%; Pred. No. 4.5;  
Matches 23; Conservative 13; Mismatches 42; Indels 0; Gaps 0;  
QY 49 EATTATEAASSDALEALADQTDALQSEAAVVKADNAASDALEALADQTDALQSEAEV 108  
Db 91 QSASSATASANSKAAKTSETNANNKNAKTSETNAASSASSASSFATAAENSARA 150  
QY 109 QSDNAASDAWEKAATPIA 126  
Db 151 SETNAGNSAQADADSKTA 168  
RESULT 19  
AK15\_RAT  
ID AK15\_RAT STANDARD; PRT; 464 AA.  
AC P24567;





CC SEQUENCE FROM N.A.  
CC MDL:LINE-94183253; PubMed-8135819;  
CC Kitajima Y., Yatsuki H., Zhang R., Matsubashi S., Hori K.;  
CC "A novel human homologue of a dead-box RNA helicase family.";  
CC Biochem. Biophys. Res. Commun. 199:748-754(1994).  
CC -1- SIMILARITY: BELONGS TO THE DEAD BOX HELICASE FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
CC EMBL: D26528; BAA05534.1; -  
CC MIM: 603793; -  
CC InterPro: IPR001410; DEAD.  
CC InterPro: IPR000629; DEAD\_ATP\_helicase.  
CC InterPro: IPR001650; Helicase\_C.  
CC Pfam: PF00270; DEAD; 1.  
CC Pfam: PF00271; helicase\_C; 1.  
CC SMART: SM00487; DEXDC; 1.  
CC SMART: SM00490; HELIC; 1.  
CC PROSITE: PS00039; DEAD\_ATP\_HELICASE; FALSE NEG.  
CC HydroLase: ATP-binding; Helicase; RNA-binding.  
CC NP\_BIND 39 46 ATP (POTENTIAL).  
CC SITE 145 148 DEAD BOX  
CC SEQUENCE 439 AA; 47405 MW; EA9E81631D5D0847 CRC64;  
Query Match 9.1%; Score 71; DB 1; Length 439;  
Best Local Similarity 26.8%; Pred. No. 2.3;  
Matches 19; Conservative 14; Mismatches 38; Indels 0; Gaps 0;  
QY 75 EEAAYKADNAASDALEALADQTDALQSEAEVQSDNAASDAWEKAATPIALDVKTKD 134  
DB 349 DDFVLKAAATFADAKQPKSTAAVKAARATAPTDQAGHQAGQAAPVTRPTTQAK 408  
QY 135 TKPVKKEERO 145  
DB 409 TVPVKPKKK 419  
RESULT 24  
SBCC\_RHOCA STANDARD; PRT; 1238 AA.  
ID SBCC\_RHOCA  
AC O68032;  
DT 01-MAR-2002 (Rel. 41, Created)  
DT 01-MAR-2002 (Rel. 41, Last sequence update)  
DE Nuclease sbcCD subunit C.  
GN SBCC.  
OS Rhodospirillum rubrum (Rhodospirillum rubrum).  
OC Bacteria; Proteobacteria; alpha subdivision; Rhodospirillum rubrum group;  
OC Rhodospirillum.  
OX NCBI\_TaxID=1061;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SB1003 / St. Louis;  
RX MDL:LINE-9740404; PubMed-9256491;  
RA Vicek C., Paces V., Maltsev N., Paces J., Haselkorn R., Fonstein M.;  
RT "Sequence of a 189-kb segment of the chromosome of Rhodospirillum  
RT capsulatus SB1003".  
RL Proc. Natl. Acad. Sci. U.S.A. 94:9384-9388(1997).  
CC -1- FUNCTION: SbcCD cleaves DNA hairpin structures. These structures  
CC can inhibit DNA replication and are intermediates in certain DNA  
CC recombination reactions. The complex acts as a 3'-5' double  
CC strand exonuclease that can open hairpins. It also has a 5'  
CC single-strand endonuclease activity (By similarity).  
CC -1- SUBUNIT: Heterodimer of sbcC and sbcD (By similarity).  
CC -1- SIMILARITY: BELONGS TO THE SMC FAMILY. SBCC SUBFAMILY.  
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
CC EMBL: AF010496; AAC16118.1; -  
CC InterPro: IPR003439; ABC\_transport.  
CC InterPro: IPR001687; ATP\_GTP\_A.  
CC HydroLase: Nuclease; Exonuclease; Endonuclease; DNA replication;  
CC DNA recombination; ATP-binding; Coiled coil.  
CC NP\_BIND 37 44 ATP (POTENTIAL).  
CC FT DOMAIN 395 438 COILED COIL (POTENTIAL).  
CC FT DOMAIN 466 487 COILED COIL (POTENTIAL).  
CC FT DOMAIN 521 600 COILED COIL (POTENTIAL).  
CC FT DOMAIN 724 770 COILED COIL (POTENTIAL).  
CC FT DOMAIN 901 943 COILED COIL (POTENTIAL).  
CC FT DOMAIN 1019 1052 COILED COIL (POTENTIAL).  
CC SQ SEQUENCE 1238 AA; 128046 MW; 2B6BC6C63A859AE1 CRC64;  
Query Match 9.1%; Score 71; DB 1; Length 1238;  
Best Local Similarity 33.3%; Pred. No. 6.7;  
Matches 20; Conservative 9; Mismatches 31; Indels 0; Gaps 0;  
QY 42 KOATDEKATTAATAAASDALEALADQTDALQSEAEVQSDNAASDALEALADQTDALQ 101  
DB 447 RRLADHRAACAQAATAAQAQAQRAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQA 506  
RESULT 25  
T2D9\_HUMAN STANDARD; PRT; 211 AA.  
ID T2D9\_HUMAN  
AC Q15344;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DE Transcription initiation factor TFIID 28 kDa subunit (TAFII-28)  
DE (TAFII28) (TFIID subunit p30-beta).  
GN TAFII OR TAF2I.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RX MDL:LINE-95246745; PubMed-7729427;  
RA Mengus G., May M., Jacq X., Staub A., Tora L., Chambon P.,  
RA Davidson I.;  
RT "Cloning and characterization of hTAFII18, hTAFII20 and hTAFII28:  
RT three subunits of the human transcription factor TFIID.";  
RL EMBO J. 14:1520-1531(1995).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MDL:LINE-96418138; PubMed-8820923;  
RA Kuzubara T., Horikoshi M.;  
RT "Isolation and characterization of a cDNA encoding a human TFIID  
RT subunit containing a variety of putative structural motifs including  
RT direct repeats.";  
RL Biol. Pharm. Bull. 19:122-126(1996).  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Williams S.;  
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: CORE TAFII PRESENT IN BOTH OF THE PREVIOUSLY DESCRIBED  
CC TFIID SPECIES WHICH EITHER LACK OR CONTAIN TAFII30 (TFIID ALPHA  
CC AND TFIID BETA RESPECTIVELY).  
CC -1- SUBUNIT: TFIID IS COMPOSED OF TATA BINDING PROTEIN (TBP) AND A  
CC NUMBER OF TBP-ASSOCIATED FACTORS (TAFs). INTERACTS WITH TAFII18  
CC BOTH IN VITRO AND INTRACELLULARLY; ALSO INTERACTS DIRECTLY WITH  
CC TBP.  
CC -1- SUBCELLULAR LOCATION: Nuclear.







```
RA Zopf D., Dineva B., Betz H., Gundelfinger E.D.;
RT *Isolation of the chicken middle-molecular weight neurofilament
RL (NF-M) gene and characterization of its promoter.*;
RN Nucleic Acids Res. 18:521-529(1990).
RP (2)
RX SEQUENCE OF 259-857 FROM N.A.
RA MEDLINE=88112814; PubMed=3123320;
RA Zopf D., Hermans-Borgmeyer I., Gundelfinger E.D., Betz H.;
RT *Identification of gene products expressed in the developing chick
RT visual system: characterization of a middle-molecular-weight
RL neurofilament cDNA.*;
RL Genes Dev. 1:699-708(1987).
CC -!- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,
CC AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
CC -!- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P, NEM IS
CC PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS
CC THOUGHT THAT PHOSPHORYLATION OF NFH RESULTS IN THE FORMATION OF
CC INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE
CC OF AXONAL CALIBER.
CC -!- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING
CC OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE
CC LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
CC COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.
CC -!- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X17102; CAA34958.1; -
CC EMBL; X05558; CAA29073.1; -
CC PIR; A27040; A27040.
CC PIR; S08061; S08061.
CC PIR; S15762; S15762.
CC InterPro: IPR001664; IF.
CC Pfam: PF00038; filament; 1.
CC PROSITE; PS00226; IF; 1.
CC Intermediate filament; Coiled coil; Neurone; Phosphorylation;
KW Glycoprotein.
FT INIT_MET 0 0
FT DOMAIN 1 98 HEAD.
FT DOMAIN 99 406 ROD.
FT DOMAIN 407 857 TAIL.
FT DOMAIN 99 130 COIL 1A.
FT DOMAIN 131 143 LINKER 1.
FT DOMAIN 144 242 COIL 1B.
FT DOMAIN 243 259 LINKER 12.
FT DOMAIN 260 281 COIL 2A.
FT DOMAIN 282 285 LINKER 2.
FT DOMAIN 286 406 COIL 2B.
FT CARBOHYD 46 46 O-LINKED (GLCNAC) (BY SIMILARITY).
FT CARBOHYD 426 426 O-LINKED (GLCNAC) (BY SIMILARITY).
FT CONFLICT 546 546 G -> R (IN REF. 2).
SQ SEQUENCE 857 AA; 95704 MW; 4E2E0FC6AC64778B CRC64;

Query Match 9.0%; Score 70; DB 1; Length 857;
Best Local Similarity 23.4%; Pred. No. 5.9;
Matches 22; Conservative 23; Mismatches 49; Indels 0; Gaps 0;

QY 30 LRNEERAIDELKKQAIEDKKAETTAIEAASSDALEALADQTDALQSEEAAYVKADNAASDA 89
Db 442 LKVOHKFVEEIEETKVESEMEDALSATAEAAQAEQEEKEAEAEAEVEEAVS 501
QY 90 LEALADQTDALQSEEAAYVQSDNAASDAWEKAAT 123
Db 502 EKAAQAAEEKEEEAEAEAEAEAEKSDAEEGS 535

RESULT 30
```

```
MST2_DROHY
ID MST2_DROHY STANDARD; PRT; 1391 AA.
AC Q08696;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Axoneme-associated protein mst101(2).
GN MST101(2).
OS Drosophila hydei (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7224;
RN [1]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RX MEDLINE=95045538; PubMed=7957199;
RA Neesen J., Padmanabhan S., Buenemann H.;
RT *Tandemly arranged repeats of a novel highly charged 16-amino-acid
RT motif representing the major component of the sperm-tail-specific
RT axoneme-associated protein family Dhmst101 form extended
RT alpha-helical rods within the extremely elongated spermatozoa of
RT Drosophila hydei.*;
RL Eur. J. Biochem. 225:1089-1095(1994).
CC -!- FUNCTION: POSSIBLE STRUCTURAL ROLE IN THE SPERM TAIL.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- TISSUE SPECIFICITY: TESTIS. PRIMARY SPERMATOCYTES AND EARLY
CC SPERMATIDS.
CC -!- DOMAIN: THE PREDOMINANT STRUCTURE IS ALPHA-HELICAL.
CC -!- POLYMORPHISM: LENGTH POLYMORPHISMS EXIST BETWEEN DIFFERENT
CC STRAINS, MOST LIKELY CAUSED BY LENGTH VARIATIONS WITHIN THE TANDEM
CC REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X73481; CAA51876.1; -
CC PIR; S34154; S34154.
CC HSSP; P01032; IG5A.
CC FlyBase: FBgn020733; Dhyd\mst101(2).
KW Sperm; Repeat; Multigene family; Polymorphism.
FT DOMAIN 332 1268 [KRI]-K-X-C-X-X-K-X-K-X-K-X-X-X-E.
FT SEQUENCE 1391 AA; 159000 MW; 1B2A368F30F48878 CRC64;

Query Match 9.0%; Score 70; DB 1; Length 1391;
Best Local Similarity 28.2%; Pred. No. 9.7;
Matches 20; Conservative 13; Mismatches 38; Indels 0; Gaps 0;

QY 22 GNAPEKALRNDEERAIDELKKQAIEDKKAETTAIEAASSDALEALADQTDALQSEEAAYVK 81
Db 753 GNNKGKKALKKEKKRELAKKAAEKKEKKAEEKKEAAEKKEKCEKTAKKRKEAEKKK 812
QY 82 ADNAASDALEA 92
Db 813 CEKTAKKRKEA 823

RESULT 31
HBHA_MYCTU
ID HBHA_MYCTU STANDARD; PRT; 198 AA.
AC Q11142; O85733;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Heparin-binding hemagglutinin (Adhesin).
GN HBHA OR RV0475 OR MT0493 OR MTCY20G9.01.
OS Mycobacterium tuberculosis, and
OS Mycobacterium bovis.
```

OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OX Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
RN NCBI\_TaxID=1773, 1765;  
[1]  
RP SEQUENCE FROM N.A., PARTIAL SEQUENCE, AND CHARACTERIZATION.  
RC SPECIES=M.tuberculosis, and M.bovis;  
RX STRAIN=H37Rv, H37Ra, and BCG / Paris 1173 P2;  
R MEDLINE=98445421; PubMed=9770536;  
RA Menozzi F.D., Bischoff R., Fort E., Brennan M.J., Loch C.;  
RT "Molecular characterization of the mycobacterial heparin-binding  
hemagglutinin, a mycobacterial adhesin."  
RL Proc. Natl. Acad. Sci. U.S.A. 95:12625-12630(1998).  
[2]  
RN SEQUENCE FROM N.A.  
RP SPECIES=M.tuberculosis; STRAIN=H37Rv;  
RX MEDLINE=98295987; PubMed=9634230;  
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,  
Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;  
RA "Deciphering the biology of Mycobacterium tuberculosis from the  
complete genome sequence."  
RL Nature 393:537-544(1998).  
[3]  
RN SEQUENCE FROM N.A.  
RP SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;  
RA Fleischmann R.D., Altand D., Eisen J.A., Carpenter L., White O.,  
Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,  
Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,  
Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
Bischoff R., Brennan M.J., Loch C.;  
RT "Identification of a heparin-binding hemagglutinin present in  
Mycobacteria."  
RL J. Exp. Med. 184:993-1001(1996).  
[5]  
RN FUNCTION.  
RP SPECIES=M.tuberculosis, and M.bovis;  
RX STRAIN=103, and BCG;  
R MEDLINE=21342355; PubMed=11449276;  
RA Pethe K., Alonso S., Biet F., Delogu G., Brennan M.J., Loch C.,  
Menozzi F.D.;  
RT "The heparin-binding haemagglutinin of M. tuberculosis is required for  
extrapulmonary dissemination."  
RL Nature 412:190-194(2001).  
[6]  
RN -1- FUNCTION: REQUIRED FOR EXTRAPULMONARY DISSEMINATION. MEDIATES  
ADHERENCE TO EPITHELIAL CELLS BY BINDING TO SULFATED  
GLYCOCONGUATES PRESENT AT THE SURFACE OF THESE CELLS; BINDS  
HEPARIN, DEXTRAN SULFATE, FUCOIDAN AND CHONDROITIN SULFATE.  
CC PROMOTES HEMAGGLUTINATION OF ERYTHROCYTES OF CERTAIN HOST SPECIES.  
CC INDUCES MYCOBACTERIAL AGGREGATION.  
CC -1- SUBCELLULAR LOCATION: SURFACE ASSOCIATED.  
CC -1- DOMAIN: HEPARIN BINDING SEEMS TO REQUIRE THE C-TERMINAL  
DOMAIN OF HBHA. PROGRESSIVE TRUNCATIONS FROM THE C-TERMINAL  
DIMINISH THE AFFINITY FOR HEPARIN.  
CC -1- PTM: GLYCOSYLATED. GLYCOSYLATION MAY PROTECT THE PROTEIN FROM  
PROTEOLYTIC DEGRADATION AND BE IMPORTANT FOR HEMAGGLUTINATION. IT  
SUGGESTS THAT THE CARBOHYDRATE MOIETY MAY BE LOCATED WITHIN THE C-  
TERMINAL DOMAIN OF HBHA.  
CC -1- MISCELLANEOUS: SERUM FROM PATIENTS DIAGNOSED WITH ACTIVE

CC TUBERCULOSIS THAT HAD NOT BEEN VACCINATED CONTAINS ANTIBODIES THAT  
CC CONTAIN AN.  
CC -1- SIMILARITY: STRONG, TO M.LEPRAE HBHA.  
CC  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: AF074390; AAC26052.1; -.  
DR EMBL: Z77162; CAB00936.1; -.  
DR EMBL: AE006951; AAK44716.1; ALT\_INIT.  
DR TIGR: MT0493; -.  
DR Tuberculin; RV0475; -.  
KW Cell adhesion; Heparin-binding; Hemagglutinin; Glycoprotein;  
KW Virulence; Complete proteome.  
FT INIT\_MET 0  
FT DOMAIN 151 193 ALA/LYS-RICH  
FT CONFLICT 120 120 R -> P (IN REF. 1).  
SQ SEQUENCE 198 AA; 21403 MW; 513760F6F1EB6042 CRC64;  
Query Match 8.9%; Score 69; DB 1; Length 198;  
Best Local Similarity 25.3%; Pred. No. 1.7;  
Matches 25; Conservative 18; Mismatches 56; Indels 0; Gaps 0;  
QY 9 RITPNGTLLNLLGNAPEKIALNEERAIDELKKQALIEDKREYTAIDAASSDALEALADQ 68  
DB 48 RVSEARLTKLQFDLPQQLTREKFTABELKRAEGYLEAATSRYNELVERGEALER 107  
QY 69 TDALOSEAAVAVADNAASDALEALADQTDALOSEAEV 107  
DB 108 LRSQSPFEVSARAEVGYDQAVELTQELGCTVASQTRAV 146  
RESULT 32  
GAR2\_SCHPO STANDARD; PRT: 500 AA.  
AC P41891; O13707;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Protein gar2  
GN GAR2 OR SPAC140.02.  
OS Schizosaccharomyces pombe (Fission yeast).  
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
OC Schizosaccharomycetes.  
OX NCBI\_TaxID=4896;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=972;  
RX MEDLINE=953119932; PubMed=7596817;  
RA Guille M.-P., Girard J.-P., Zabetakis D., Lapeyre B., Melese T.,  
Caizerques-Ferrer M.;  
RT "gar2 is a nucleolar protein from Schizosaccharomyces pombe required  
for 18S rRNA and 40S ribosomal subunit accumulation."  
RL Nucleic Acids Res. 23:1912-1918(1995).  
[2]  
RN SEQUENCE FROM N.A.  
RC STRAIN=972;  
RA Brown D., Churcher C.M., Barrell B.G., Rajandream M.A., Wood V.;  
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: HELPS THE ASSEMBLY OF PRE-RIBOSOMAL PARTICLES  
CONTAINING 18S rRNA.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR; NUCLEOLUS.  
CC -1- SIMILARITY: BELONGS TO THE GAR FAMILY.  
CC -1- SIMILARITY: CONTAINS 2 RNA RECOGNITION MOTIFS (RRM).  
CC  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration

**THIS PAGE BLANK (USPTO)**

; LENGTH: 16135  
; TYPE: DNA  
; ORGANISM: DROSOPHILA  
US-09-614-150-35057

Alignment Scores:  
Pred. No.: 2.29e+03 Length: 16135  
Score: 52.00 Matches: 11  
Percent Similarity: 65.22% Conservative: 4  
Best Local Similarity: 47.83% Mismatches: 8  
Query Match: 40.31% Indels: 0  
DB: 23 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-09-614-150-35057 (1-16135)

Qy 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVallys 23  
Db 5971 CTGGGAAGTCTCTTCAAGCAGGTGGAGTCTCTCAAGCCGCGAGTTGCTCTCTGTGGCA 5912  
Qy 24 AlaAspAsn 26  
Db 5911 GCCGACAA 5903

RESULT 40  
US-09-619-049-1340/c  
; Sequence 1340, Application US/09619049  
; GENERAL INFORMATION:  
; APPLICANT: YANDELL, MARK  
; TITLE OF INVENTION: ISOLATED DROSOPHILA PROTEINS ESSENTIAL  
; TITLE OF INVENTION: FOR SURVIVAL, NUCLEIC ACID MOLECULES ENCODING ESSENTIAL  
; TITLE OF INVENTION: DROSOPHILA PROTEINS, AND USES THEREOF AS INSECTICIDAL  
; TITLE OF INVENTION: TARGETS  
; FILE REFERENCE: CL000735  
; CURRENT APPLICATION NUMBER: US/09/619,049  
; CURRENT FILING DATE: 2000-07-18  
; PRIOR APPLICATION NUMBER: 60/171,590  
; PRIOR FILING DATE: 1999-12-23  
; PRIOR APPLICATION NUMBER: 60/171,627  
; PRIOR FILING DATE: 1999-12-23  
; PRIOR APPLICATION NUMBER: 60/175,763  
; PRIOR FILING DATE: 2000-01-12  
; PRIOR APPLICATION NUMBER: 60/175,685  
; PRIOR FILING DATE: 2000-01-12  
; PRIOR APPLICATION NUMBER: 60/186,663  
; PRIOR FILING DATE: 2000-03-03  
; PRIOR APPLICATION NUMBER: 60/187,241  
; PRIOR FILING DATE: 2000-03-03  
; NUMBER OF SEQ ID NOS: 1533  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1340  
; LENGTH: 16135  
; TYPE: DNA  
; ORGANISM: DROSOPHILA  
US-09-619-049-1340

Alignment Scores:  
Pred. No.: 2.29e+03 Length: 16135  
Score: 52.00 Matches: 11  
Percent Similarity: 65.22% Conservative: 4  
Best Local Similarity: 47.83% Mismatches: 8  
Query Match: 40.31% Indels: 0  
DB: 23 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-09-619-049-1340 (1-16135)

Qy 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVallys 23  
Db 5971 CTGGGAAGTCTCTTCAAGCAGGTGGAGTCTCTCAAGCCGCGAGTTGCTCTCTGTGGCA 5912  
Qy 24 AlaAspAsn 26  
Db 5911 GCCGACAA 5903

Search completed: October 13, 2002, 04:38:41  
Job time : 564.612 secs

ORGANISM: Arabidopsis thaliana  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)-(1230)  
OTHER INFORMATION: any n = a, g, c, t, unknown, or other  
NAME/KEY: misc\_feature  
LOCATION: (1)-(1230)  
OTHER INFORMATION: 40182 : FINISHED (Clone Number : FINISHED)  
US-09-565-309A-62025

Alignment Scores:  
Pred. No.: 116 Length: 1230  
Score: 52.00 Matches: 11  
Percent Similarity: 72.73% Conservatives: 5  
Best Local Similarity: 50.00% Mismatches: 6  
Query Match: 40.31% Indels: 0  
DB: 22 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-09-565-309A-62025 (1-1230)

Qy 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVallys 23  
||| |||||:||||: ||| ||| :||||: |||||: ||| |||  
Db 518 CTTCGGCAGCTCTCGGACAAAGGACATGCTCAATTCAAAAGTGCATCGCAAG 459

Qy 24 AlaAsp 25  
|||||  
Db 458 GCGGAT 453

RESULT 37

US-09-595-329A-545/c  
Sequence 545, Application US/09595329A

GENERAL INFORMATION:

APPLICANT: ALEXANDROV, Nikolai

APPLICANT: VYACHESLAV, Brover

TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptide

TITLE OF INVENTION: Thereby

FILE REFERENCE: 2750-0948P

CURRENT APPLICATION NUMBER: US/09/595,329A

CURRENT FILING DATE: 2000-06-16

NUMBER OF SEQ ID NOS: 3309

SOFTWARE: PatentIn version 3.0

SEQ ID NO 545

LENGTH: 1230

TYPE: DNA

ORGANISM: Arabidopsis Thaliana

FEATURE:

NAME/KEY: misc\_feature

LOCATION: (1)-(1230)

OTHER INFORMATION: Ceres Seq. ID no. 1013068

NAME/KEY: misc\_feature

LOCATION: (1)-(1230)

OTHER INFORMATION: n is a, c, t, g, unknown, or other

US-09-595-329A-545

Alignment Scores:  
Pred. No.: 116 Length: 1230  
Score: 52.00 Matches: 11  
Percent Similarity: 72.73% Conservatives: 5  
Best Local Similarity: 50.00% Mismatches: 6  
Query Match: 40.31% Indels: 0  
DB: 22 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-09-595-329A-545 (1-1230)

Qy 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVallys 23  
||| |||||:||||: ||| ||| :||||: |||||: ||| |||  
Db 518 CTTCGGCAGCTCTCGGACAAAGGACATGCTCAATTCAAAAGTGCATCGCAAG 459

Qy 24 AlaAsp 25  
|||||  
Db 458 GCGGAT 453

RESULT 38

US-09-565-309A-52656/c  
Sequence 52656, Application US/09565309A  
GENERAL INFORMATION:  
APPLICANT: ALEXANDROV, Nikolai  
APPLICANT: BROVER, Vyacheslav  
TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTID  
TITLE OF INVENTION: THEREBY  
FILE REFERENCE: 2750-0853P  
CURRENT APPLICATION NUMBER: US/09/565,309A  
CURRENT FILING DATE: 2000-05-05  
NUMBER OF SEQ ID NOS: 68449  
SEQ ID NO 52656  
LENGTH: 1244  
TYPE: DNA  
ORGANISM: Arabidopsis thaliana  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)-(1244)  
OTHER INFORMATION: any n = a, g, c, t, unknown, or other  
NAME/KEY: misc\_feature  
LOCATION: (1)-(1244)  
OTHER INFORMATION: 40182 : 5TAG CONSENSUS (Clone Number:5tag\_consensus)  
US-09-565-309A-52656

Alignment Scores:  
Pred. No.: 117 Length: 1244  
Score: 52.00 Matches: 11  
Percent Similarity: 72.73% Conservatives: 5  
Best Local Similarity: 50.00% Mismatches: 6  
Query Match: 40.31% Indels: 0  
DB: 22 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-09-565-309A-52656 (1-1244)

Qy 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVallys 23  
||| |||||:||||: ||| ||| :||||: |||||: ||| |||  
Db 532 CTTCGGCAGCTCTCGGACAAAGGACATGCTCAATTCAAAAGTGCATCGCAAG 473

Qy 24 AlaAsp 25  
|||||  
Db 472 GCGGAT 467

RESULT 39

US-09-614-150-35057/c

Sequence 35057, Application US/09614150

GENERAL INFORMATION:

APPLICANT: Venter, J. Craig

APPLICANT: et al.

TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID

TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE

TITLE OF INVENTION: DROSOPHILA GENES.

FILE REFERENCE: CL000728

CURRENT APPLICATION NUMBER: US/09/614,150

CURRENT FILING DATE: 2000-07-11

PRIOR APPLICATION NUMBER: 60/157,832

PRIOR FILING DATE: 1999-10-05

PRIOR APPLICATION NUMBER: 60/160,191

PRIOR FILING DATE: 1999-10-19

PRIOR APPLICATION NUMBER: 60/161,932

PRIOR FILING DATE: 1999-10-28

PRIOR APPLICATION NUMBER: 60/164,769

PRIOR FILING DATE: 1999-11-12

PRIOR APPLICATION NUMBER: 60/173,383

PRIOR FILING DATE: 1999-12-28

PRIOR APPLICATION NUMBER: 60/175,693

PRIOR FILING DATE: 2000-01-12

PRIOR APPLICATION NUMBER: 60/184,831

PRIOR FILING DATE: 2000-02-24

PRIOR APPLICATION NUMBER: 60/191,637

PRIOR FILING DATE: 2000-03-23

NUMBER OF SEQ ID NOS: 43008

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 35057

Db 401 GCGGAT 396

RESULT 33

US-09-565-309A-67148/c

; Sequence 67148, Application US/09565309A

; GENERAL INFORMATION:

; APPLICANT: ALEXANDROV, Nikolai

; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES

; TITLE OF INVENTION: THEREBY

; FILE REFERENCE: 2750-0853P

; CURRENT APPLICATION NUMBER: US/09/565,309A

; CURRENT FILING DATE: 2000-05-05

; NUMBER OF SEQ ID NOS: 68449

; SEQ ID NO 67148

; LENGTH: 939

; TYPE: DNA

; ORGANISM: Arabidopsis thaliana

; FEATURE:

; NAME/KEY: misc\_feature

; LOCATION: (1)..(939)

; OTHER INFORMATION: any n = a, g, c, t, unknown, or other

; NAME/KEY: misc\_feature

; LOCATION: (1)..(939)

; OTHER INFORMATION: 40182 : CDS (Clone Number : CDS)

US-09-565-309A-67148

Alignment Scores:

Pred. No.:	84.7	Length:	939
Score:	52.00	Matches:	11
Percent Similarity:	72.73%	Conservative:	5
Best Local Similarity:	50.00%	Mismatches:	6
Query Match:	40.31%	Indels:	0
DB:	22	Gaps:	0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-09-565-309A-67148 (1-939)

QY 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVallys 23

Db 382 CTTCCGCGCACTCTCCGACAAAGGACATGCCCTCAAAATTCAAAAGCTGCAATCGCAAG 323

QY 24 AlaAsp 25

Db 322 GCGGAT 317

RESULT 34

US-09-708-427-29368/c

; Sequence 29368, Application US/09708427

; GENERAL INFORMATION:

; APPLICANT: N. ALEXANDROV et al.

; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES

; TITLE OF INVENTION: THEREBY

; FILE REFERENCE: 2750-1243P

; CURRENT APPLICATION NUMBER: US/09/708,427

; CURRENT FILING DATE: 2000-11-09

; NUMBER OF SEQ ID NOS: 85364

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 29368

; LENGTH: 984

; TYPE: DNA

; ORGANISM: Arabidopsis thaliana

; FEATURE:

; NAME/KEY: misc\_feature

; LOCATION: 1..984

; OTHER INFORMATION: any n = a, g, c, t, unknown, or other

; NAME/KEY: misc\_feature

; LOCATION: 1..984

; OTHER INFORMATION: Ceres Seq. ID 1825505

US-09-708-427-29368

Alignment Scores:

Pred. No.:	89.5	Length:	984
Score:	52.00	Matches:	11

Percent Similarity: 72.73% Conservative: 5

Best Local Similarity: 50.00% Mismatches: 6

Query Match: 40.31% Indels: 0

DB: 28 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-09-708-427-29368 (1-984)

QY 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVallys 23

Db 442 CTTCCGCGCACTCTCCGACAAAGGACATGCCCTCAAAATTCAAAAGCTGCAATCGCAAG 383

QY 24 AlaAsp 25

Db 382 GCGGAT 377

RESULT 35

US-09-565-309A-67147/c

; Sequence 67147, Application US/09565309A

; GENERAL INFORMATION:

; APPLICANT: ALEXANDROV, Nikolai

; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTID

; TITLE OF INVENTION: THEREBY

; FILE REFERENCE: 2750-0853P

; CURRENT APPLICATION NUMBER: US/09/565,309A

; CURRENT FILING DATE: 2000-05-05

; NUMBER OF SEQ ID NOS: 68449

; SEQ ID NO 67147

; LENGTH: 999

; TYPE: DNA

; ORGANISM: Arabidopsis thaliana

; FEATURE:

; NAME/KEY: misc\_feature

; LOCATION: (1)..(999)

; OTHER INFORMATION: any n = a, g, c, t, unknown, or other

; NAME/KEY: misc\_feature

; LOCATION: (1)..(999)

; OTHER INFORMATION: 40182 : CDS (Clone Number : CDS)

US-09-565-309A-67147

Alignment Scores:

Pred. No.:	91	Length:	999
Score:	52.00	Matches:	11
Percent Similarity:	72.73%	Conservative:	5
Best Local Similarity:	50.00%	Mismatches:	6
Query Match:	40.31%	Indels:	0
DB:	22	Gaps:	0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-09-565-309A-67147 (1-999)

QY 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVallys 23

Db 442 CTTCCGCGCACTCTCCGACAAAGGACATGCCCTCAAAATTCAAAAGCTGCAATCGCAAG 383

QY 24 AlaAsp 25

Db 382 GCGGAT 377

RESULT 36

US-09-565-309A-62025/c

; Sequence 62025, Application US/09565309A

; GENERAL INFORMATION:

; APPLICANT: ALEXANDROV, Nikolai

; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTID

; TITLE OF INVENTION: THEREBY

; FILE REFERENCE: 2750-0853P

; CURRENT APPLICATION NUMBER: US/09/565,309A

; CURRENT FILING DATE: 2000-05-05

; NUMBER OF SEQ ID NOS: 68449

; SEQ ID NO 62025

; LENGTH: 1230

; TYPE: DNA

```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17589
; LENGTH: 7206
; TYPE: DNA
; ORGANISM: Drosophila
US-60-167-217-17589

Alignment Scores:
Pred. No.: 402          Length: 7206
Score: 54.00           Matches: 12
Percent Similarity: 66.67% Conservative: 4
Best Local Similarity: 50.00% Mismatches: 8
Query Match: 41.86%      Indels: 0
DB: 55                 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x US-60-167-217-17589 (1-7206)
Qy 5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVallysAla 24
Db 4905 GAGGAGGAGCTGACGAGGAGCGCCGCGAGGAGGATGCGCGCGCC 4964
Qy 25 AspAsnAlaAla 28
Db 4965 GATGCTGCGCGC 4976

RESULT 30
US-09-528-237A-1811/c
; Sequence 1811, Application US/09528237A
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; TITLE OF INVENTION: The Drosophila Genome, Primary Nucleic
; TITLE OF INVENTION: Acid Sequences, Systems Containing The Nucleic Acid
; FILE REFERENCE: C1000284
; CURRENT APPLICATION NUMBER: US/09/528,237A
; CURRENT FILING DATE: 2000-03-17
; NUMBER OF SEQ ID NOS: 2926
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1811
; LENGTH: 32275
; TYPE: DNA
; ORGANISM: Drosophila
US-09-528-237A-1811

Alignment Scores:
Pred. No.: 2,29e+03     Length: 32275
Score: 54.00           Matches: 12
Percent Similarity: 66.67% Conservative: 4
Best Local Similarity: 50.00% Mismatches: 8
Query Match: 41.86%      Indels: 0
DB: 19                 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x US-09-528-237A-1811 (1-32275)
Qy 5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVallysAla 24
Db 26995 GAGGAGGAGCTGACGAGGAGCGCCGCGAGGAGGATGCGCGCGCC 26936
Qy 25 AspAsnAlaAla 28
Db 26935 GATGCTGCGCGC 26924

RESULT 31
US-09-565-309A-32821/c
; Sequence 32821, Application US/09565309A
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nickolai
; APPLICANT: BROVER, Vyacheslav
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; TITLE OF INVENTION: THEREBY
; FILE REFERENCE: 2750-0853P
; CURRENT APPLICATION NUMBER: US/09/565,309A
; CURRENT FILING DATE: 2000-05-05
```

```
; NUMBER OF SEQ ID NOS: 68449
; SEQ ID NO 32821
; LENGTH: 594
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(594)
; OTHER INFORMATION: any n = a, g, c, t, unknown, or other
; NAME/KEY: misc_feature
; LOCATION: (1)..(594)
; OTHER INFORMATION: 40182:107311 (Clone Number:Unique Sequence Identifier)
US-09-565-309A-32821

Alignment Scores:
Pred. No.: 49.8         Length: 594
Score: 52.00           Matches: 11
Percent Similarity: 72.73% Conservative: 5
Best Local Similarity: 50.00% Mismatches: 6
Query Match: 40.31%      Indels: 0
DB: 22                 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x US-09-565-309A-32821 (1-594)
Qy 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVallys 23
Db 518 CTTCCGCGCACTCTCCGACAAAGGAGGACATGCCTCAAAATTCAAAAAGCTGCAATCGCAAG 459
Qy 24 AlaAsp 25
Db 458 GCGGAT 453

RESULT 32
US-09-565-309A-32824/c
; Sequence 32824, Application US/09565309A
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nickolai
; APPLICANT: BROVER, Vyacheslav
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTID
; TITLE OF INVENTION: THEREBY
; FILE REFERENCE: 2750-0853P
; CURRENT APPLICATION NUMBER: US/09/565,309A
; CURRENT FILING DATE: 2000-05-05
; NUMBER OF SEQ ID NOS: 68449
; SEQ ID NO 32824
; LENGTH: 641
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(641)
; OTHER INFORMATION: any n = a, g, c, t, unknown, or other
; NAME/KEY: misc_feature
; LOCATION: (1)..(641)
; OTHER INFORMATION: 40182:957281 (Clone Number:Unique Sequence Identifier)
US-09-565-309A-32824

Alignment Scores:
Pred. No.: 54.4         Length: 641
Score: 52.00           Matches: 11
Percent Similarity: 72.73% Conservative: 5
Best Local Similarity: 50.00% Mismatches: 6
Query Match: 40.31%      Indels: 0
DB: 22                 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x US-09-565-309A-32824 (1-641)
Qy 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVallys 23
Db 461 CTTCCGCGCACTCTCCGACAAAGGAGGACATGCCTCAAAATTCAAAAAGCTGCAATCGCAAG 402
Qy 24 AlaAsp 25
Db 458 GCGGAT 453
```

```
; LENGTH: 6398
; TYPE: DNA
; ORGANISM: DROSOPHILA
US-09-614-150-17530

Alignment Scores:
Pred. No.: 350 Length: 6398
Score: 54.00 Matches: 12
Percent Similarity: 66.67% Conservative: 4
Best Local Similarity: 50.00% Mismatches: 8
Query Match: 41.86% Indels: 0
DB: 23 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x US-09-614-150-17530 (1-6398)
QY 5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluAlaAlaValValLysAla 24
   ||| |||||: ||||| :|||: |||||
Db 2298 GAGGAGGAAGCTGACGAGGAGGAGCGCGCGGACGAGGATGCCGCGGCTGCCGCC 2239
QY 25 AspAsnAlaAla 28
   ||| |||||
Db 2238 GATGCTGCGCGC 2227

RESULT 26
US-60-191-637-17580/c
; Sequence 17580, Application US/60191637
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; TITLE OF INVENTION: NUCLEIC ACID DETECTION KITS COMPRISING
; TITLE OF INVENTION: GENE SEQUENCES EXPRESSED FROM THE DROSOPHILA GENOME, AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL000392
; CURRENT APPLICATION NUMBER: US/60/191,637
; CURRENT FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 42660
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17580
; LENGTH: 6398
; TYPE: DNA
; ORGANISM: DROSOPHILA
US-60-191-637-17580

Alignment Scores:
Pred. No.: 350 Length: 6398
Score: 54.00 Matches: 12
Percent Similarity: 66.67% Conservative: 4
Best Local Similarity: 50.00% Mismatches: 8
Query Match: 41.86% Indels: 0
DB: 58 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x US-60-191-637-17580 (1-6398)
QY 5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluAlaAlaValValLysAla 24
   ||| |||||: ||||| :|||: |||||
Db 2298 GAGGAGGAAGCTGACGAGGAGGAGCGCGCGGACGAGGATGCCGCGGCTGCCGCC 2239
QY 25 AspAsnAlaAla 28
   ||| |||||
Db 2238 GATGCTGCGCGC 2227

RESULT 27
US-60-191-681-13896/c
; Sequence 13896, Application US/60191681
; GENERAL INFORMATION:
; APPLICANT: Li, Peter, W.D.
; TITLE OF INVENTION: ISOLATED DROSOPHILA PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING ISOLATED DROSOPHILA PROTEINS AND
; TITLE OF INVENTION: USES THEREOF.
; FILE REFERENCE: c1000390
; CURRENT APPLICATION NUMBER: US/60/191,681
; CURRENT FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 30973
; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 13896
; LENGTH: 6398
; TYPE: DNA
; ORGANISM: DROSOPHILA
US-60-191-681-13896

Alignment Scores:
Pred. No.: 350 Length: 6398
Score: 54.00 Matches: 12
Percent Similarity: 66.67% Conservative: 4
Best Local Similarity: 50.00% Mismatches: 8
Query Match: 41.86% Indels: 0
DB: 58 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x US-60-191-681-13896 (1-6398)
QY 5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluAlaAlaValValLysAla 24
   ||| |||||: ||||| :|||: |||||
Db 2298 GAGGAGGAAGCTGACGAGGAGGAGCGCGCGGACGAGGATGCCGCGGCTGCCGCC 2239
QY 25 AspAsnAlaAla 28
   ||| |||||
Db 2238 GATGCTGCGCGC 4976

RESULT 29
US-60-167-217-17589
; Sequence 17589, Application US/60167217
; GENERAL INFORMATION:
; APPLICANT: Li, Peter W. D.
; TITLE OF INVENTION: ISOLATED DROSOPHILA PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING DROSOPHILA PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL000152
; CURRENT APPLICATION NUMBER: US/60/167,217
; CURRENT FILING DATE: 1999-11-24
; NUMBER OF SEQ ID NOS: 23195
```



```
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17531
; LENGTH: 3349
; TYPE: DNA
; ORGANISM: DROSOPHILA
US-09-614-150-17531
```

```
Alignment Scores:
Pred. No.: 165 Length: 3349
Score: 54.00 Matches: 12
Percent Similarity: 66.67% Conservative: 4
Best Local Similarity: 50.00% Mismatches: 8
Query Match: 41.86% Indels: 0
DB: 23 Gaps: 0
```

US-09-847-539a-6\_COPY\_59\_86 (1-28) x US-09-614-150-17531 (1-3349)

```
QY 5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVallysAla 24
   ||| |||||:  ||||| :|||:||||| |||
Db 1122 GAGGAGGAAGCTGACGAGGAGCGCCGCCGAGGAGGATGCCCGCGCTGCCGCC 1063

QY 25 AspAsnAlaAla 28
   ||| |||||
Db 1062 GATGCTGCGCGC 1051
```

RESULT 23

US-60-191-637-17581/c

```
; Sequence 17581, Application US/60191637
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; TITLE OF INVENTION: NUCLEIC ACID DETECTION KITS COMPRISING
; TITLE OF INVENTION: GENE SEQUENCES EXPRESSED FROM THE DROSOPHILA GENOME, AND
; FILE REFERENCE: CL000392
; CURRENT APPLICATION NUMBER: US/60/191,637
; NUMBER OF SEQ ID NOS: 42660
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17581
; LENGTH: 3349
; TYPE: DNA
; ORGANISM: DROSOPHILA
US-60-191-637-17581
```

```
Alignment Scores:
Pred. No.: 165 Length: 3349
Score: 54.00 Matches: 12
Percent Similarity: 66.67% Conservative: 4
Best Local Similarity: 50.00% Mismatches: 8
Query Match: 41.86% Indels: 0
DB: 23 Gaps: 0
```

US-09-847-539a-6\_COPY\_59\_86 (1-28) x US-60-191-637-17581 (1-3349)

```
QY 5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVallysAla 24
   ||| |||||:  ||||| :|||:||||| |||
Db 1122 GAGGAGGAAGCTGACGAGGAGCGCCGCCGAGGAGGATGCCCGCGCTGCCGCC 1063
```

^

```
QY 25 AspAsnAlaAla 28
   ||| |||||
Db 1062 GATGCTGCGCGC 1051
```

RESULT 24

```
US-60-191-681-13897/c
; Sequence 13897, Application US/60191681
; GENERAL INFORMATION:
; APPLICANT: Li, Peter, W.D.
; TITLE OF INVENTION: ISOLATED DROSOPHILA PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING ISOLATED DROSOPHILA PROTEINS AND
; TITLE OF INVENTION: USES THEREOF.
; FILE REFERENCE: C1000390
; CURRENT APPLICATION NUMBER: US/60/191,681
; NUMBER OF SEQ ID NOS: 30973
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13897
; LENGTH: 3349
; TYPE: DNA
; ORGANISM: DROSOPHILA
US-60-191-681-13897
```

```
Alignment Scores:
Pred. No.: 165 Length: 3349
Score: 54.00 Matches: 12
Percent Similarity: 66.67% Conservative: 4
Best Local Similarity: 50.00% Mismatches: 8
Query Match: 41.86% Indels: 0
DB: 23 Gaps: 0
```

US-09-847-539a-6\_COPY\_59\_86 (1-28) x US-60-191-681-13897 (1-3349)

```
QY 5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVallysAla 24
   ||| |||||:  ||||| :|||:||||| |||
Db 1122 GAGGAGGAAGCTGACGAGGAGCGCCGCCGAGGAGGATGCCCGCGCTGCCGCC 1063

QY 25 AspAsnAlaAla 28
   ||| |||||
Db 1062 GATGCTGCGCGC 1051
```

RESULT 25

US-09-614-150-17530/c

```
; Sequence 17530, Application US/09614150
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/09/614,150
; CURRENT FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17530
```

```
; LENGTH: 1783
; TYPE: DNA
; ORGANISM: Drosophila
US-60-142-845-289

Alignment Scores:
Pred. No.: 79.6      Length: 1783
Score: 54.00      Matches: 12
Percent Similarity: 66.67%      Conservative: 4
Best Local Similarity: 50.00%      Mismatches: 8
Query Match: 41.86%      Indels: 0
DB: 53      Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x US-60-142-845-289 (1-1783)

Qy 5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLysAla 24
   ||| |||||: ||||| :|||: ||||| |||
Db 1475 GAGGAGGAAGCTGACGAGGAGGACGCCGCCGAGGATGCGCGGCTGCCGCC 1534

Qy 25 AsnAlaAla 28
   ||| |||||
Db 1535 GATGCTCGCGCG 1546

RESULT 19
US-60-145-134-143
; Sequence 143, Application US/60145134
; GENERAL INFORMATION:
; APPLICANT: Bonazzi, Vivien
; TITLE OF INVENTION: ISOLATED DROSOPHILA PROTEINS, NUCLEIC
; FILE REFERENCE: CL000058
; CURRENT APPLICATION NUMBER: US/60/145,134
; CURRENT FILING DATE: 1999-07-22
; NUMBER OF SEQ ID NOS: 342
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 143
; LENGTH: 1785
; TYPE: DNA
; ORGANISM: Drosophila
US-60-145-134-143

Alignment Scores:
Pred. No.: 79.7      Length: 1785
Score: 54.00      Matches: 12
Percent Similarity: 66.67%      Conservative: 4
Best Local Similarity: 50.00%      Mismatches: 8
Query Match: 41.86%      Indels: 0
DB: 53      Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x US-60-145-134-143 (1-1785)

Qy 5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLysAla 24
   ||| |||||: ||||| :|||: ||||| |||
Db 1477 GAGGAGGAAGCTGACGAGGAGGACGCCGCCGAGGATGCGCGGCTGCCGCC 1536

Qy 25 AsnAlaAla 28
   ||| |||||
Db 1537 GATGCTCGCGCG 1548

RESULT 20
US-60-173-464-14401/c
; Sequence 14401, Application US/60173464
; GENERAL INFORMATION:
; APPLICANT: Li, Peter W.D.
; TITLE OF INVENTION: ISOLATED G-PROTEIN COUPLED RECEPTORS,
; FILE REFERENCE: CL000173
; CURRENT APPLICATION NUMBER: US/60/173,464
; CURRENT FILING DATE: 1999-12-29
; NUMBER OF SEQ ID NOS: 30269
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14401
```

```
; LENGTH: 3046
; TYPE: DNA
; ORGANISM: Drosophila
US-60-173-464-14401

Alignment Scores:
Pred. No.: 148      Length: 3046
Score: 54.00      Matches: 12
Percent Similarity: 66.67%      Conservative: 4
Best Local Similarity: 50.00%      Mismatches: 8
Query Match: 41.86%      Indels: 0
DB: 56      Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x US-60-173-464-14401 (1-3046)

Qy 5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLysAla 24
   ||| |||||: ||||| :|||: ||||| |||
Db 1014 GAGGAGGAAGCTGACGAGGAGGACGCCGCCGAGGATGCGCGGCTGCCGCC 955

Qy 25 AsnAlaAla 28
   ||| |||||
Db 954 GATGCTCGCGCG 943

RESULT 21
US-60-167-217-17590/c
; Sequence 17590, Application US/60167217
; GENERAL INFORMATION:
; APPLICANT: Li, Peter W. D.
; TITLE OF INVENTION: ISOLATED DROSOPHILA PROTEINS, NUCLEIC
; FILE REFERENCE: CL000152
; CURRENT APPLICATION NUMBER: US/60/167,217
; CURRENT FILING DATE: 1999-11-24
; NUMBER OF SEQ ID NOS: 23195
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17590
; LENGTH: 3049
; TYPE: DNA
; ORGANISM: Drosophila
US-60-167-217-17590

Alignment Scores:
Pred. No.: 148      Length: 3049
Score: 54.00      Matches: 12
Percent Similarity: 66.67%      Conservative: 4
Best Local Similarity: 50.00%      Mismatches: 8
Query Match: 41.86%      Indels: 0
DB: 55      Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x US-60-167-217-17590 (1-3049)

Qy 5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLysAla 24
   ||| |||||: ||||| :|||: ||||| |||
Db 1017 GAGGAGGAAGCTGACGAGGAGGACGCCGCCGAGGATGCGCGGCTGCCGCC 958

Qy 25 AsnAlaAla 28
   ||| |||||
Db 957 GATGCTCGCGCG 946

RESULT 22
US-09-614-150-17531/c
; Sequence 17531, Application US/09614150
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/09/614,150
; CURRENT FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/157,832
```

```
* ; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 10031
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: Windows NT 4.0
; SOFTWARE: UNIX
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/993.002A
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: GTN-018
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)742-4214
; INFORMATION FOR SEQ ID NO: 4235:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1239 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1...1239
US-08-993-002A-4235
Alignment Scores:
Pred. No.: 15.6 Length: 1239
Score: 57.00 Matches: 12
Percent Similarity: 69.57% Conservative: 4
Best Local Similarity: 52.17% Mismatches: 7
Query Match: 44.19% Indels: 0
DB: 13 Gaps: 0
US-09-847-539A-6_COPY_59_86 (1-28) x US-08-993-002A-4235 (1-1239)
Qy 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVallys 23
Db 145 ATACAGGCCCTACAGGACCAATTGCCTTTAGATTCTCAAGAAAAAGTCGTTAGCAAA 204
Qy 24 AlaAspAsn 26
Db 205 TGGGATAAC 213
RESULT 17
US-08-621-425-49
; Sequence 49, Application US/08621425
; GENERAL INFORMATION:
; APPLICANT: Douglas Smith
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; TITLE OF INVENTION: RELATING TO HELICOBACTER PYLORI FOR
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 495
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street, Suite 510
```

```
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 8mm cartridge tape
; COMPUTER: SPARC station LX
; OPERATING SYSTEM: sunOS4
; SOFTWARE: tar
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/621.425
; FILING DATE: 25-MARCH-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/561,469
; FILING DATE: 17-NOV-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/487,032
; FILING DATE: 07-JUNE-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: GTN-001CP2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 49:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3497 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1...3497
US-08-621-425-49
Alignment Scores:
Pred. No.: 51.9 Length: 3497
Score: 57.00 Matches: 12
Percent Similarity: 69.57% Conservative: 4
Best Local Similarity: 52.17% Mismatches: 7
Query Match: 44.19% Indels: 0
DB: 10 Gaps: 0
US-09-847-539A-6_COPY_59_86 (1-28) x US-08-621-425-49 (1-3497)
Qy 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVallys 23
Db 1528 ATACAGGCCCTACAGGACCAATTGCCTTTAGATTCTCAAGAAAAAGTCGTTAGCAAA 1587
Qy 24 AlaAspAsn 26
Db 1588 TGGGATAAC 1596
RESULT 18
US-60-142-845-289
; Sequence 289, Application US/60142845
; GENERAL INFORMATION:
; APPLICANT: Kerlavage, Anthony
; TITLE OF INVENTION: ISOLATED DROSOPHILA PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING DROSOPHILA PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CLO00048
; CURRENT APPLICATION NUMBER: US/60/142.845
; CURRENT FILING DATE: 1999-07-08
; NUMBER OF SEQ ID NOS: 704
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 289
```

```

; MEDIUM TYPE:
; COMPUTER:
; OPERATING SYSTEM:
; SOFTWARE:
; CURRENT APPLICATION DATA: PCT/US97/19575
; FILING DATE:
; PRIOR APPLICATION DATA:
; FILING DATE: 28-OCT-1996
; APPLICATION NUMBER: US 08/739,150
; FILING DATE: 28-OCT-1996
; PRIOR APPLICATION DATA:
; FILING DATE: 06-DEC-1996
; APPLICATION NUMBER: US 08/759,739
; FILING DATE: 17-NOV-1995
; PRIOR APPLICATION DATA:
; FILING DATE: 01-APRIL-1996
; APPLICATION NUMBER: US 08/891,928
; FILING DATE: 14-JULY-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: GTN-001CP10PC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 51:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1239 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1...1239
PCT-US97-19575-51

```

```

Alignment Scores:
Pred. No.: 15.6 Length: 1239
Score: 57.00 Matches: 12
Percent Similarity: 69.57% Conservative: 4
Best Local Similarity: 52.17% Mismatches: 7
Query Match: 44.19% Indels: 0
DB: 1 Gaps: 0

```

US-09-847-539a-6\_COPY\_59\_86 (1-28) x PCT-US97-19575-51 (1-1239)

```

Qy 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVallys 23
Db 145 ATACAAGCCCTACAGGAGCAAAATGACGCTTTAGATTCTCAAGAAAAGTCGTTAGCAAA 204
Qy 24 AlaAspAsn 26
Db 205 TGGGATAAC 213

```

```

RESULT 15
US-08-759-739-250
; Sequence 250, Application US/08759739
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS SMITH ET AL
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; TITLE OF INVENTION: RELATING TO HELICOBACTER PYLORI VACCINE COMPOSITIONS THERE
; NUMBER OF SEQUENCES: 608
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street, Suite 510
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:

```

```

; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER:
; OPERATING SYSTEM:
; SOFTWARE:
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/759,739
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/487,032
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/561,469
; FILING DATE: 17-NOV-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/630,405
; FILING DATE: 01-APRIL-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/660,742
; FILING DATE: 06-JUN-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/736,791
; FILING DATE: 25-OCT-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/739,150
; FILING DATE: 28-OCT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: GTN-001CP8
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 250:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1239 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1...1239
US-08-759-739-250

```

```

Alignment Scores:
Pred. No.: 15.6 Length: 1239
Score: 57.00 Matches: 12
Percent Similarity: 69.57% Conservative: 4
Best Local Similarity: 52.17% Mismatches: 7
Query Match: 44.19% Indels: 0
DB: 11 Gaps: 0

```

US-09-847-539a-6\_COPY\_59\_86 (1-28) x US-08-759-739-250 (1-1239)

```

Qy 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVallys 23
Db 145 ATACAAGCCCTACAGGAGCAAAATGACGCTTTAGATTCTCAAGAAAAGTCGTTAGCAAA 204
Qy 24 AlaAspAsn 26
Db 205 TGGGATAAC 213

```

```

RESULT 16
US-08-993-002A-4235
; Sequence 4235, Application US/08993002A
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS SMITH et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; TITLE OF INVENTION: RELATING TO HELICOBACTER PYLORI FOR

```

```

; APPLICATION NUMBER: US/08/759,739
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/487,032
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/561,469
; FILING DATE: 17-NOV-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/630,405
; FILING DATE: 01-APRIL-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/660,742
; FILING DATE: 06-JUN-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/736,791
; FILING DATE: 25-OCT-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/739,150
; FILING DATE: 28-OCT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: GTN-001CP8
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 92:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 666 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1...666
US-08-759-739-92
Alignment Scores:
Pred. No.: 7.59 Length: 666
Score: 57.00 Matches: 12
Percent Similarity: 69.57% Conservative: 4
Best Local Similarity: 52.17% Mismatches: 7
Query Match: 44.19% Indels: 0
DB: 11 Gaps: 0
US-09-847-539A-6_COPY_59_86 (1-28) x US-08-759-739-92 (1-666)
QY 4 LeuGluaLeuAlaaspGlnThrAspAlaLeuGlnSerGluGluAlaValValLys 23
Db 229 ATACAAGCCCTACAGGAGCAAAATTGAGCTTTAGATTCTCAAGAAAAAGTCGTAGCAAA 288
QY 24 AlaAspAsn 26
Db 289 TGGGATAAC 297
RESULT 13
; Sequence 4234, Application US/08993002A
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS SMITH et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; TITLE OF INVENTION: RELATING TO HELICOBACTER PYLORI FOR
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 10031
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
APPLICATION NUMBER: US/08/993,002A
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Mandragouras, Amy E.
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: GTN-018
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)742-4214
INFORMATION FOR SEQ ID NO: 4234:
SEQUENCE CHARACTERISTICS:
LENGTH: 666 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Helicobacter pylori
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1...666
US-08-993-002A-4234
Alignment Scores:
Pred. No.: 7.59 Length: 666
Score: 57.00 Matches: 12
Percent Similarity: 69.57% Conservative: 4
Best Local Similarity: 52.17% Mismatches: 7
Query Match: 44.19% Indels: 0
DB: 13 Gaps: 0
US-09-847-539A-6_COPY_59_86 (1-28) x US-08-993-002A-4234 (1-666)
QY 4 LeuGluaLeuAlaaspGlnThrAspAlaLeuGlnSerGluGluAlaValValLys 23
Db 229 ATACAAGCCCTACAGGAGCAAAATTGAGCTTTAGATTCTCAAGAAAAAGTCGTAGCAAA 288
QY 24 AlaAspAsn 26
Db 289 TGGGATAAC 297
RESULT 14
; Sequence 51, Application PC/TUS9719575
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; TITLE OF INVENTION: RELATING TO HELICOBACTER PYLORI AND
; TITLE OF INVENTION: VACCINE COMPOSITIONS THEREOF
; NUMBER OF SEQUENCES: 208
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:

```

```
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/561,469
; FILING DATE: 17-NOV-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/630,405
; FILING DATE: 01-APRIL-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/660,742
; FILING DATE: 06-JUN-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/736,791
; FILING DATE: 25-OCT-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/739,150
; FILING DATE: 28-OCT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: GTN-001CP8
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 43:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 576 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1...576
US-08-759-739-43

Alignment Scores:
Pred. No.: 6.41 Length: 576
Score: 57.00 Matches: 12
Percent Similarity: 69.57% Conservative: 4
Best Local Similarity: 52.17% Mismatches: 7
Query Match: 44.19% Indels: 0
DB: 11 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x US-08-759-739-43 (1-576)
Qy 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVallys 23
Db 145 ATACAGCCCTACAGGACAAATTGACGCTTTAGATTCTCAAGAAAAAGTCGTTAGCAA 204

Qy 24 AlaAspAsn 26
Db 205 TGGGATAAC 213

RESULT 11
US-08-993-002A-4232
; Sequence 4232, Application US/08993002A
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS SMITH et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; TITLE OF INVENTION: RELATING TO HELICOBACTER PYLORI FOR
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 10031
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
```

```
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: Windows NT 4.0
; SOFTWARE: UNIX
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/993.002A
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: GTN-018
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)742-4214
; INFORMATION FOR SEQ ID NO: 4232:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 576 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1...576
US-08-993-002A-4232

Alignment Scores:
Pred. No.: 6.41 Length: 576
Score: 57.00 Matches: 12
Percent Similarity: 69.57% Conservative: 4
Best Local Similarity: 52.17% Mismatches: 7
Query Match: 44.19% Indels: 0
DB: 13 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x US-08-993-002A-4232 (1-576)
Qy 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVallys 23
Db 145 ATACAGCCCTACAGGACAAATTGACGCTTTAGATTCTCAAGAAAAAGTCGTTAGCAA 204

Qy 24 AlaAspAsn 26
Db 205 TGGGATAAC 213

RESULT 12
US-08-759-739-92
; Sequence 92, Application US/08759739
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS SMITH ET AL
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; TITLE OF INVENTION: RELATING TO HELICOBACTER PYLORI VACCINE COMPOSITIONS THERE
; NUMBER OF SEQUENCES: 608
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street, Suite 510
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER:
; OPERATING SYSTEM:
; SOFTWARE:
; CURRENT APPLICATION DATA:
```



; APPLICANT: Rassmussen, Magnus  
; TITLE OF INVENTION: STREPTOCOCCAL ALPHA 2M BINDING PROTEIN  
; FILE REFERENCE: 100084.415US / N.75312B  
; CURRENT APPLICATION NUMBER: US/09/847,539A  
; CURRENT FILING DATE: 2001-03-01  
; NUMBER OF SEQ ID NOS: 30  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 27  
; LENGTH: 764  
; TYPE: DNA  
; ORGANISM: Streptococcus pyogenes  
US-09-847-539A-27

Alignment Scores:  
Pred. No.: 2,25e-12 Length: 764  
Score: 129.00 Matches: 28  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 32 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-09-847-539A-27 (1-764)

QY 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluAlaAla 20  
|||||  
Db 333 TCAGATGCCCTTAGAGCATTAGCGGATCAACAGACGCTTTACAAATCAGAAGAAGCTGGC 392  
QY 21 ValValLysAlaAspAsnAlaAla 28  
|||||  
Db 393 GTTGTAAAGCGGATAACGCTGCT 416

## RESULT 5

US-09-847-539A-13  
; Sequence 13, Application US/09847539A  
; GENERAL INFORMATION:  
; APPLICANT: Bjorck, Lars H  
; TITLE OF INVENTION: STREPTOCOCCAL ALPHA 2M BINDING PROTEIN  
; FILE REFERENCE: 100084.415US / N.75312B  
; CURRENT APPLICATION NUMBER: US/09/847,539A  
; CURRENT FILING DATE: 2001-03-01  
; NUMBER OF SEQ ID NOS: 30  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 13  
; LENGTH: 777  
; TYPE: DNA  
; ORGANISM: Streptococcus pyogenes  
US-09-847-539A-13

Alignment Scores:  
Pred. No.: 2,29e-12 Length: 777  
Score: 129.00 Matches: 28  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 32 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-09-847-539A-13 (1-777)

QY 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluAlaAla 20  
|||||  
Db 319 ACTGACGCCTTAGAGCATTGCGGATCAACAGACGCTTTACAAATCAGAAGAAGCTGGC 378  
QY 21 ValValLysAlaAspAsnAlaAla 28  
|||||  
Db 379 GTTGTAAAGCGGATAACGCTGCT 402

## RESULT 6

US-09-847-539A-15  
; Sequence 15, Application US/09847539A  
; GENERAL INFORMATION:  
; APPLICANT: Bjorck, Lars H  
; APPLICANT: Rassmussen, Magnus

; TITLE OF INVENTION: STREPTOCOCCAL ALPHA 2M BINDING PROTEIN  
; FILE REFERENCE: 100084.415US / N.75312B  
; CURRENT APPLICATION NUMBER: US/09/847,539A  
; CURRENT FILING DATE: 2001-03-01  
; NUMBER OF SEQ ID NOS: 30  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 15  
; LENGTH: 853  
; TYPE: DNA  
; ORGANISM: Streptococcus pyogenes  
US-09-847-539A-15

Alignment Scores:  
Pred. No.: 2,56e-12 Length: 853  
Score: 129.00 Matches: 28  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 32 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-09-847-539A-15 (1-853)

QY 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluAlaAla 20  
|||||  
Db 556 ACTGACGCCTTAGAGCATTGCGGATCAACAGACGCTTTACAAATCAGAAGAAGCTGGC 615  
QY 21 ValValLysAlaAspAsnAlaAla 28  
|||||  
Db 616 GTTGTAAAGCGGATAACGCTGCT 639

## RESULT 7

US-09-489-039A-318  
; Sequence 318, Application US/09489039A  
; GENERAL INFORMATION:  
; APPLICANT: Gary Breton et. al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
; FILE REFERENCE: 2709.2004001  
; CURRENT APPLICATION NUMBER: US/09/489,039A  
; CURRENT FILING DATE: 2000-01-27  
; PRIOR APPLICATION NUMBER: US 60/117,747  
; PRIOR FILING DATE: 1999-01-29  
; NUMBER OF SEQ ID NOS: 14342  
; SEQ ID NO 318  
; LENGTH: 828  
; TYPE: DNA  
; ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-318

Alignment Scores:  
Pred. No.: 6.53 Length: 828  
Score: 58.00 Matches: 12  
Percent Similarity: 75.00% Conservative: 6  
Best Local Similarity: 50.00% Mismatches: 6  
Query Match: 44.96% Indels: 0  
DB: 18 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-09-489-039A-318 (1-828)

QY 2 AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluAlaAlaVal 21  
|||||  
Db 157 GACCGTCTCAGCGGCTCGCGACACGAGCTGCAGGCGCGCTACGCCATCCAGCAAGCATC 216  
QY 22 ValLysAlaAsp 25  
|||||  
Db 217 CTCAAAGCGGAT 228

## RESULT 8

US-08-487-032A-282  
; Sequence 282, Application US/08487032A  
; GENERAL INFORMATION:  
; APPLICANT: DOUGLAS SMITH  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES



```
17 57 44.2 3497 10 US-08-621-425-49 Sequence 49, Appl
18 54 41.9 1783 53 US-60-142-845-289 Sequence 289, App
19 54 41.9 1785 53 US-60-145-134-143 Sequence 143, App
20 54 41.9 3046 56 US-60-173-464-14401 Sequence 14401, A
21 54 41.9 3049 55 US-60-167-217-17590 Sequence 17590, A
22 54 41.9 3349 23 US-09-614-150-17531 Sequence 17531, A
23 54 41.9 3349 58 US-60-191-637-17581 Sequence 17581, A
24 54 41.9 3349 58 US-60-191-681-13897 Sequence 13897, A
25 54 41.9 6398 23 US-09-614-150-17530 Sequence 17530, A
26 54 41.9 6398 58 US-60-191-637-17580 Sequence 17580, A
27 54 41.9 6398 58 US-60-191-681-13896 Sequence 13896, A
28 54 41.9 7203 56 US-60-173-464-14400 Sequence 14400, A
29 54 41.9 7206 55 US-60-167-217-17589 Sequence 17589, A
30 54 41.9 32275 19 US-09-528-237A-1811 Sequence 1811, Ap
31 52 40.3 594 22 US-09-565-309A-32821 Sequence 32821, A
32 52 40.3 641 22 US-09-565-309A-32824 Sequence 32824, A
33 52 40.3 939 22 US-09-565-309A-67148 Sequence 67148, A
34 52 40.3 984 28 US-09-708-427-29368 Sequence 29368, A
35 52 40.3 999 22 US-09-565-309A-67147 Sequence 67147, A
36 52 40.3 1230 22 US-09-565-309A-62025 Sequence 62025, A
37 52 40.3 1230 22 US-09-595-329A-545 Sequence 545, App
38 52 40.3 1244 22 US-09-565-309A-52656 Sequence 52656, A
39 52 40.3 16135 23 US-09-614-150-35057 Sequence 35057, A
40 52 40.3 16135 23 US-09-619-049-1340 Sequence 1340, Ap
41 52 40.3 16135 56 US-60-171-627-1946 Sequence 1946, Ap
42 52 40.3 16135 56 US-60-173-464-26803 Sequence 26803, A
43 52 40.3 16135 58 US-60-191-637-34656 Sequence 34656, A
44 52 40.3 16135 58 US-60-191-681-27255 Sequence 27255, A
45 52 40.3 18349 56 US-60-171-627-1945 Sequence 1945, Ap
```

## ALIGNMENTS

```
RESULT 1
US-09-847-539A-14
; Sequence 14, Application US/09847539A
; GENERAL INFORMATION:
; APPLICANT: Bjorck, Lars H
; APPLICANT: Rasmussen, Magnus
; TITLE OF INVENTION: STREPTOCOCCAL ALPHA ZM BINDING PROTEIN
; FILE REFERENCE: 100084.415US / N.75312B
; CURRENT APPLICATION NUMBER: US/09/847,539A
; CURRENT FILING DATE: 2001-05-01
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
; LENGTH: 469
; TYPE: DNA
; ORGANISM: Streptococcus pyogenes
US-09-847-539A-14
```

```
Alignment Scores:
Pred. No.: 1-28e-12 Length: 469
Score: 129.00 Matches: 28
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 32 Gaps: 0
```

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-09-847-539A-14 (1-469)

```
QY 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaA 20
|||||
Db 172 TCAGATGCCTTAGAAGCATTAGCGGATCAACAGACGCTTTACAATCAGAAGAAGCTGCG 231
|||||
QY 21 ValValLysAlaAspAsnAlaA 28
|||||
Db 232 GTTGTAAAGCGGATAACGCTGCT 255
|||||
```

```
RESULT 2
US-09-847-539A-16
; Sequence 16, Application US/09847539A
; GENERAL INFORMATION:
```

```
; APPLICANT: Bjorck, Lars H
; APPLICANT: Rasmussen, Magnus
; TITLE OF INVENTION: STREPTOCOCCAL ALPHA ZM BINDING PROTEIN
; FILE REFERENCE: 100084.415US / N.75312B
; CURRENT APPLICATION NUMBER: US/09/847,539A
; CURRENT FILING DATE: 2001-05-01
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 16
; LENGTH: 504
; TYPE: DNA
; ORGANISM: Streptococcus pyogenes
US-09-847-539A-16
```

```
Alignment Scores:
Pred. No.: 1-39e-12 Length: 504
Score: 129.00 Matches: 28
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 32 Gaps: 0
```

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-09-847-539A-16 (1-504)

```
QY 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaA 20
|||||
Db 208 TCAGATGCCTTAGAAGCATTAGCGGATCAACAGACGCTTTACAATCAGAAGAAGCTGCG 267
|||||
QY 21 ValValLysAlaAspAsnAlaA 28
|||||
Db 268 GTTGTAAAGCGGATAACGCTGCT 291
|||||
```

```
RESULT 3
US-09-847-539A-12
; Sequence 12, Application US/09847539A
; GENERAL INFORMATION:
; APPLICANT: Bjorck, Lars H
; APPLICANT: Rasmussen, Magnus
; TITLE OF INVENTION: STREPTOCOCCAL ALPHA ZM BINDING PROTEIN
; FILE REFERENCE: 100084.415US / N.75312B
; CURRENT APPLICATION NUMBER: US/09/847,539A
; CURRENT FILING DATE: 2001-05-01
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 12
; LENGTH: 654
; TYPE: DNA
; ORGANISM: Streptococcus pyogenes
US-09-847-539A-12
```

```
Alignment Scores:
Pred. No.: 1-88e-12 Length: 654
Score: 129.00 Matches: 28
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 32 Gaps: 0
```

US-09-847-539A-6\_COPY\_59\_86 (1-28) x US-09-847-539A-12 (1-654)

```
QY 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaA 20
|||||
Db 274 TCAGATGCCTTAGAAGCATTAGCGGATCAACAGACGCTTTACAATCAGAAGAAGCTGCG 333
|||||
QY 21 ValValLysAlaAspAsnAlaA 28
|||||
Db 334 GTTGTAAAGCGGATAACGCTGCT 357
|||||
```

```
RESULT 4
US-09-847-539A-27
; Sequence 27, Application US/09847539A
; GENERAL INFORMATION:
; APPLICANT: Bjorck, Lars H
```

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 13, 2002, 02:20:12 : Search time 553.112 Seconds:  
(without alignments)  
1095.215 Million cell updates/sec

Title: US-09-847-539A-6\_COPY\_59\_86  
Perfect score: 129  
Sequence: 1 SDALEALADOTDALOSEAAVVKADNAA 28

Scoring table: BLOSUM62  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 21979536 seqs, 10817449327 residues

Total number of hits satisfying chosen parameters: 43959072

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters: ~  
-MODEL=frame\_plus\_p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool/US09847539/runat\_10102002\_092549\_3486/app\_query.fasta\_1.526  
-DB=Pending\_Patents\_NA\_Main -QFMT=fastap -SUFFIX=trmp -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blos62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=40  
-MODE=LOCAL -OUTFMT=ptlo -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USPR=US09847539 -CGN\_1\_1201\_@runa\_10102002\_092549\_3486 -NCPU=6 -ICPU=3  
-NO\_XLPHY -NO\_MAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : Pending\_Patents\_NA\_Main:  
1: /cgn2\_6/ptodata/2/pna/PCTUS\_COMB.seq.\*  
2: /cgn2\_6/ptodata/2/pna/US06\_COMB.seq.\*  
3: /cgn2\_6/ptodata/2/pna/US07\_COMB.seq.\*  
4: /cgn2\_6/ptodata/2/pna/US08\_COMB.seq.\*  
5: /cgn2\_6/ptodata/2/pna/US081\_COMB.seq.\*  
6: /cgn2\_6/ptodata/2/pna/US082\_COMB.seq.\*  
7: /cgn2\_6/ptodata/2/pna/US083\_COMB.seq.\*  
8: /cgn2\_6/ptodata/2/pna/US084\_COMB.seq.\*  
9: /cgn2\_6/ptodata/2/pna/US085\_COMB.seq.\*  
10: /cgn2\_6/ptodata/2/pna/US086\_COMB.seq.\*  
11: /cgn2\_6/ptodata/2/pna/US087\_COMB.seq.\*  
12: /cgn2\_6/ptodata/2/pna/US088\_COMB.seq.\*  
13: /cgn2\_6/ptodata/2/pna/US089\_COMB.seq.\*  
14: /cgn2\_6/ptodata/2/pna/US090\_COMB.seq.\*  
15: /cgn2\_6/ptodata/2/pna/US091\_COMB.seq.\*  
16: /cgn2\_6/ptodata/2/pna/US092\_COMB.seq.\*  
17: /cgn2\_6/ptodata/2/pna/US093\_COMB.seq.\*  
18: /cgn2\_6/ptodata/2/pna/US094\_COMB.seq.\*  
19: /cgn2\_6/ptodata/2/pna/US095A\_COMB.seq.\*  
20: /cgn2\_6/ptodata/2/pna/US095B\_COMB.seq.\*  
21: /cgn2\_6/ptodata/2/pna/US095C\_COMB.seq.\*  
22: /cgn2\_6/ptodata/2/pna/US095D\_COMB.seq.\*  
23: /cgn2\_6/ptodata/2/pna/US096A\_COMB.seq.\*  
24: /cgn2\_6/ptodata/2/pna/US096B\_COMB.seq.\*  
25: /cgn2\_6/ptodata/2/pna/US096C\_COMB.seq.\*  
26: /cgn2\_6/ptodata/2/pna/US096D\_COMB.seq.\*  
27: /cgn2\_6/ptodata/2/pna/US096E\_COMB.seq.\*  
28: /cgn2\_6/ptodata/2/pna/US097A\_COMB.seq.\*

29: /cgn2\_6/ptodata/2/pna/US097B\_COMB.seq.\*  
30: /cgn2\_6/ptodata/2/pna/US097C\_COMB.seq.\*  
31: /cgn2\_6/ptodata/2/pna/US098A\_COMB.seq.\*  
32: /cgn2\_6/ptodata/2/pna/US098B\_COMB.seq.\*  
33: /cgn2\_6/ptodata/2/pna/US098C\_COMB.seq.\*  
34: /cgn2\_6/ptodata/2/pna/US099A\_COMB.seq.\*  
35: /cgn2\_6/ptodata/2/pna/US099B\_COMB.seq.\*  
36: /cgn2\_6/ptodata/2/pna/US099C\_COMB.seq.\*  
37: /cgn2\_6/ptodata/2/pna/US100\_COMB.seq.\*  
38: /cgn2\_6/ptodata/2/pna/US101\_COMB.seq.\*  
39: /cgn2\_6/ptodata/2/pna/US6000\_COMB.seq.\*  
40: /cgn2\_6/ptodata/2/pna/US6001\_COMB.seq.\*  
41: /cgn2\_6/ptodata/2/pna/US6002\_COMB.seq.\*  
42: /cgn2\_6/ptodata/2/pna/US6003\_COMB.seq.\*  
43: /cgn2\_6/ptodata/2/pna/US6004\_COMB.seq.\*  
44: /cgn2\_6/ptodata/2/pna/US6005\_COMB.seq.\*  
45: /cgn2\_6/ptodata/2/pna/US6006\_COMB.seq.\*  
46: /cgn2\_6/ptodata/2/pna/US6007\_COMB.seq.\*  
47: /cgn2\_6/ptodata/2/pna/US6008\_COMB.seq.\*  
48: /cgn2\_6/ptodata/2/pna/US6009\_COMB.seq.\*  
49: /cgn2\_6/ptodata/2/pna/US6010\_COMB.seq.\*  
50: /cgn2\_6/ptodata/2/pna/US6011\_COMB.seq.\*  
51: /cgn2\_6/ptodata/2/pna/US6012\_COMB.seq.\*  
52: /cgn2\_6/ptodata/2/pna/US6013\_COMB.seq.\*  
53: /cgn2\_6/ptodata/2/pna/US6014\_COMB.seq.\*  
54: /cgn2\_6/ptodata/2/pna/US6015\_COMB.seq.\*  
55: /cgn2\_6/ptodata/2/pna/US6016\_COMB.seq.\*  
56: /cgn2\_6/ptodata/2/pna/US6017\_COMB.seq.\*  
57: /cgn2\_6/ptodata/2/pna/US6018\_COMB.seq.\*  
58: /cgn2\_6/ptodata/2/pna/US6019\_COMB.seq.\*  
59: /cgn2\_6/ptodata/2/pna/US6020\_COMB.seq.\*  
60: /cgn2\_6/ptodata/2/pna/US6021\_COMB.seq.\*  
61: /cgn2\_6/ptodata/2/pna/US6022\_COMB.seq.\*  
62: /cgn2\_6/ptodata/2/pna/US6023\_COMB.seq.\*  
63: /cgn2\_6/ptodata/2/pna/US6024\_COMB.seq.\*  
64: /cgn2\_6/ptodata/2/pna/US6025\_COMB.seq.\*  
65: /cgn2\_6/ptodata/2/pna/US6026\_COMB.seq.\*  
66: /cgn2\_6/ptodata/2/pna/US6027\_COMB.seq.\*  
67: /cgn2\_6/ptodata/2/pna/US6028\_COMB.seq.\*  
68: /cgn2\_6/ptodata/2/pna/US6029\_COMB.seq.\*  
69: /cgn2\_6/ptodata/2/pna/US6030\_COMB.seq.\*  
70: /cgn2\_6/ptodata/2/pna/US6031\_COMB.seq.\*  
71: /cgn2\_6/ptodata/2/pna/US6032\_COMB.seq.\*  
72: /cgn2\_6/ptodata/2/pna/US6033\_COMB.seq.\*  
73: /cgn2\_6/ptodata/2/pna/US6034\_COMB.seq.\*  
74: /cgn2\_6/ptodata/2/pna/US6035\_COMB.seq.\*  
75: /cgn2\_6/ptodata/2/pna/US6036\_COMB.seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	129	100.0	469	32	US-09-847-539A-14	Sequence 14, Appl
2	129	100.0	504	32	US-09-847-539A-16	Sequence 16, Appl
3	129	100.0	654	32	US-09-847-539A-12	Sequence 12, Appl
4	129	100.0	764	32	US-09-847-539A-27	Sequence 27, Appl
5	129	100.0	777	32	US-09-847-539A-13	Sequence 43, Appl
6	129	100.0	853	32	US-09-847-539A-15	Sequence 15, Appl
7	58	45.0	848	18	US-09-489-039A-318	Sequence 318, App
8	57	44.2	576	8	US-08-487-032A-282	Sequence 282, App
9	57	44.2	576	9	US-08-561-469A-282	Sequence 282, App
10	57	44.2	576	11	US-08-759-739-43	Sequence 43, Appl
11	57	44.2	576	13	US-08-993-002A-4232	Sequence 4232, Ap
12	57	44.2	666	11	US-08-759-739-92	Sequence 92, Appl
13	57	44.2	666	13	US-08-993-002A-4234	Sequence 4234, Ap
14	57	44.2	1239	11	PCT-US97-19575-51	Sequence 51, Appl
15	57	44.2	1239	11	US-08-759-739-250	Sequence 250, App
16	57	44.2	1239	13	US-08-993-002A-4235	Sequence 4235, Ap

```

REFERENCE 2 (bases 1 to 848)
AUTHORS Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
          Bolotin-Fukuhara,M., Bon,E., Bottier,P., Casaregola,S.,
          de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
          Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
          Saurin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
          Wincker,P. and Weissenbach,J.
TITLE Genomic Exploration of the Hemiascomycetous Yeasts: 1. A set of
JOURNAL yeast species for molecular evolution studies(1)
PUBMED FEBS Lett. 487 (1), 3-12 (2000)
AUTHORS 11152876
JOURNAL 3 (bases 1 to 848)
        Genoscope.
        Direct Submission
        Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,
        2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
        seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT This STS is part of a random genomic sequencing program of thirteen
        yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
        exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
        Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
        lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
        angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
        Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
        5 kb were prepared and both extremities were sequenced. See
        keywords for description of this sequence and for the sequence of
        the other extremity of this insert.
FEATURES             Location/Qualifiers
     source           1..848
                     /organism="Saccharomyces bayanus"
                     /strain="CUB 533"
                     /variety="uvarum"
                     /db_xref="taxon:4931"
                     /clone="AS0AA018G01"
                     /clone_lib="AS0AA"
                     /note="end : 17"
     misc_feature     <20..>180
                     /note="similar to Saccharomyces cerevisiae ORF YOL110w [
                     SHR5 ; RAS suppressor ]
                     1 putative frameshift(s)"
                     /evidence=not_experimental
     misc_feature     <577..>846
                     /note="similar to Saccharomyces cerevisiae ORF YOL109w [
                     ZE01 ; zeocin resistance ]"
                     /evidence=not_experimental
BASE COUNT 211 a 205 c 189 g 240 t
ORIGIN
Alignment Scores:
Pred. No.: 0.0504 Length: 848
Score: 88.00 Matches: 21
Percent Similarity: 58.82% Conservative: 9
Best Local Similarity: 41.18% Mismatches: 21
Query Match: 11.33% Indels: 0
DB: 11 Gaps: 0
US-09-847-539A-6 (1-159) x CNS0618W (1-848)
QY 81 LysAlaAspAsnAlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeu 100
DB 595 AAAGCTGAAACCTGCGCTGAAGACCTCCAAACAAAGCTGGATGAACCAAGGAGTCCTTA 654
QY 101 GlnSerGluAlaGluValValGlnSerAspAsnAlaAlaSerAspAlaTrpGluLys 120
DB 655 CAACACAGAGCGCGAGCTGAAGGACACACAGCCCAACTGCTGTCGACGACTTGAAGAG 714
QY 121 AlaAlaThrProIleAlaLeuAspValLysLys 131
DB 715 GAAGCTACCCAGAGCCGCAACAAAGTTAAGAAG 747

```

Search completed: October 13, 2002, 03:00:40  
 Job time : 1854.74 secs













```

Oy 106 GluValValGlnSerAspAsnAlaSerAspAlaTrpGluLysAlaAlaThrProIle 125
    ||| :::: ||||| ::||| ::||| |||||
Db 123840 GAGGAGCGAGCGCCCAAGAGCGCGGAGAGCTGCCCAAAAGGTGCGGAAGAGCC 123781
Oy 126 AlaLeu 127
    |||||
Db 123780 GCTTTA 123775

RESULT 34
LOCUS AC020883 269726 bp DNA linear HTG 15-JUL-2000
DEFINITION Mus musculus clone RP23-441L2, WORKING DRAFT SEQUENCE, 56 unordered
pieces.
ACCESSION AC020883
VERSION AC020883.3 GI:9211213
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 269726)
TITLE DOE Joint Genome Institute.
JOURNAL Sequencing of Mouse
REFERENCE 2 (bases 1 to 269726)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (10-JAN-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Jul 15, 2000 this sequence version replaced gi:6980210.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 1906482
Center clone name: RPCI-23_441L2
-----
Summary Statistics
Consensus quality: 205453 bases at least Q40
Consensus quality: 224960 bases at least Q30
Consensus quality: 223624 bases at least Q20
Estimated insert size: 240000; pulse field gel estimation
Estimated insert size: 264226; sum-of-contigs estimation
Quality coverage: 8.93 in Q20 bases; pulse field gel estimation
Quality coverage: 8.11 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 56 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1018: contig of 1018 bp in length
* 1019: gap of unknown length
* 1119: contig of 1030 bp in length
* 2149: gap of unknown length
* 3292: contig of 1044 bp in length
* 3293: gap of unknown length
* 3393: contig of 1687 bp in length
* 5080: gap of unknown length
* 5180: contig of 1196 bp in length
* 6376: gap of unknown length
* 6476: contig of 1307 bp in length
* 7783: gap of unknown length
* 7883: contig of 1098 bp in length
* 8981: gap of unknown length
* 9081: contig of 1110 bp in length
* 10191: gap of unknown length
* 10291: contig of 1368 bp in length
* 11659: gap of unknown length
* 11659: contig of 1367 bp in length

11759 12977: contig of 1219 bp in length
13077: gap of unknown length
13978 12977: contig of 1167 bp in length
14244: gap of unknown length
14345 12977: contig of 1526 bp in length
14345 12977: gap of unknown length
15870 12977: gap of unknown length
15871 12977: contig of 1004 bp in length
16974 12977: gap of unknown length
17074 12977: contig of 1025 bp in length
18099 12977: gap of unknown length
18199 12977: contig of 1430 bp in length
19629 12977: gap of unknown length
19630 12977: contig of 1845 bp in length
21574 12977: gap of unknown length
21674 12977: contig of 1203 bp in length
22877 12977: gap of unknown length
22878 12977: contig of 1173 bp in length
24150 12977: gap of unknown length
24151 12977: contig of 1196 bp in length
25446 12977: gap of unknown length
25447 12977: contig of 2278 bp in length
27824 12977: gap of unknown length
27825 12977: contig of 1019 bp in length
28943 12977: gap of unknown length
28944 12977: contig of 1147 bp in length
30190 12977: gap of unknown length
30191 12977: contig of 2020 bp in length
32310 12977: gap of unknown length
32311 12977: contig of 1540 bp in length
33950 12977: gap of unknown length
34050 12977: contig of 1118 bp in length
35168 12977: gap of unknown length
35268 12977: contig of 1326 bp in length
36594 12977: gap of unknown length
36694 12977: contig of 1271 bp in length
37965 12977: gap of unknown length
37966 12977: contig of 1144 bp in length
39209 12977: gap of unknown length
39210 12977: contig of 1040 bp in length
40349 12977: gap of unknown length
40449 12977: contig of 1165 bp in length
41614 12977: gap of unknown length
41714 12977: contig of 1309 bp in length
43023 12977: gap of unknown length
43123 12977: contig of 1631 bp in length
44754 12977: gap of unknown length
44854 12977: contig of 1976 bp in length
46830 12977: gap of unknown length
46930 12977: contig of 1266 bp in length
48196 12977: gap of unknown length
48296 12977: contig of 1178 bp in length
49474 12977: gap of unknown length
49574 12977: contig of 1070 bp in length
50644 12977: gap of unknown length
50744 12977: contig of 1072 bp in length
51816 12977: gap of unknown length
51916 12977: contig of 1322 bp in length
53238 12977: gap of unknown length
53338 12977: contig of 1485 bp in length
54823 12977: gap of unknown length
54923 12977: contig of 2134 bp in length
57057 12977: gap of unknown length
57157 12977: contig of 1226 bp in length
58383 12977: gap of unknown length
58483 12977: contig of 1758 bp in length
60241 12977: gap of unknown length
60341 12977: contig of 1226 bp in length
61567 12977: gap of unknown length
61667 12977: contig of 1620 bp in length
63287 12977: gap of unknown length
63387 12977: contig of 1489 bp in length
64876 12977: gap of unknown length
64976 12977: contig of 1367 bp in length
66343 12977: contig of 1367 bp in length

```

```
/protein_id="AAG22139.1"
/db_xref="GI:10726439"
/translation="MATSKWESDEEIOYFREYLRIPIVSHPDYAPCVFELRRQANLM
DLPKMYYPANEQNPVVLTKGLNPELPSILLNSHMDVVPFENWTHPPFGADIDE
EGRIFAKQDMKSGMQLAAVRAALKRSGAKFRKTHISEVADDEMGGRYGMRRFPVP
TEDRALNVGFMDEGLASPEQIPLFYAERAVRVYFNISCTAGHSGLLPNTAGEK
LNYTVGRMMERQRQVRLQNNPELVIGDVTINLTKLGGVQSNVVPFLMWCFDCR
LADVDPEERANLHKCADVGGIEITYEOKPKVPPTADGSPNFWLAFKKATDEM
HISIKPQIFGCGTDSRYIRAVGIPALGFSPNNTPVLLHDHDEFIQADIYLRGVQIFQ
KIISNIANV"
complement(join(<32971. .33155,33431. .34088,34241. .>34261))
/gene="CG14688"
/product="CT34474"
/db_xref="FLYBASE:FBan0014688"
/db_xref="FBgn0037819"
/evidence=not_experimental
complement(<32971. .>34261)
/gene="CG14688"
/map="86C2-86C2"
/db_xref="FLYBASE:FBan0014688"
/db_xref="FLYBASE:FBgn0037819"
/evidence=not_experimental
complement(join(32971. .33155,33431. .34088,34241. .34261))
/gene="CG14688"
/codon_start=1
/db_xref="FLYBASE:FBan0014688"
/db_xref="FBgn0037819"
/evidence=not_experimental
/protein_id="AAF54534.2"
/db_xref="GI:10726440"
/translation="MQSNRSLNENGYIVIEDFLTAEBVDLSYQNGRALCLDAPQNNR
KIFSTIQEDAQGRLEFYMESGDKVRYPFEKAGVGDQQLLVDMIALNKVGHALHVE
HPANATFSNRVREIQWLNYNPAVCQSMYIYKNPGVGVEVPHODSWFLHTDPNS
AVFWLALEDTLQNGCLOFTKGSHKSGHRRYLRNPKDKASELMVYDRAAPIYPOSS
PTPMQVSKGTCILIHGNVHKSEPNRSQKSHAYTFHVETENNKYSEDNWLQAPKD
KPFVFLFERKA"
join(42080. .42386,44345. .44848,45327. .>45631)
/gene="CG14681"
/product="CT34467"
/db_xref="FLYBASE:FBan0014681"
/db_xref="FBgn0037820"
<42080. .>45631
/gene="CG14681"
/map="86C2-86C2"
/db_xref="FLYBASE:FBan0014681"
/db_xref="FBgn0037820"
join(42326. .42386,44345. .44848,45327. .45631)
/gene="CG14681"
/note="CG14681 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0014681"
/db_xref="FBgn0037820"
/protein_id="AAF54535.1"
/db_xref="GI:7299343"
/translation="MLAKMKPWLILLCLLAALSCLASFGDAAPYYVYTKIGALTRLH
HGVSQDYYVDSRTIFIKFNKYNDGEAPAYFYVGNRTARPSNEGAARLDERGGTASLT
RVRNKKDVTLSLPGKTLRDIKWFSVWCDEFAVNFQDVSTIPNLDFPRPKISALRGV
HGVSSDNIYVDAQTLVLPNPSYDGEAPDAKFWYGRQRTSDGLRIPDENGKFNPLR
RYERKTVLTLPELDTIFDIGHFGWCEAFVDFGHVRLPEGLNVPVPSLKMIGTSPOV
KKNASQILSELOR"
join(<46218. .46232,46434. .46533,47430. .48206,48265. .48458,
48515. .>51130)
/gene="CG14682"
/product="CT34468"
/db_xref="FLYBASE:FBan0014682"
/db_xref="FBgn0037821"
/evidence=not_experimental
<46218. .>51130
/gene="CG14682"
/map="86C2-86C2"
/db_xref="FLYBASE:FBan0014682"
/db_xref="FBgn0037821"

/proofread=1
/translation="MOTFRSLNCEVLYDDLAFAEVRVAVAGESIVVQLVAKLEPNHIYM
SEGISPNKNIISOMIGADAVAAVNDPOTNGFATDYFLEGKQCSGGRGACPDTKISEK
TNSIRLLNAAWNGYSIVYORSLAATRLDLPISITCAESVVAWALGPLNDYQEVSFII
TFYNKHLHQIEFGROPKWNCPLEPGARGNSSEDEDSAPAAQSTGGAGYPPAGRPN
VEPDEFEYENRAEALHRQPPORQETAITQRRPVPTPKPVNSNGAMDIPAICQHEPE
DGVFYAQMGPTGGKHGYPATICHVGWISWYINGLLIPELIHVVRGKTYTFVVEGNNP
DIPAKYHPFYISDDPVGGYEHKREKKAIRYAGVHRSRQVPTCTGVCRCNWKTPD
VEGPADDYQSGGAYORTLTCLKDAGEPVTWKPDRTPTVYVICTHRYLGNKIIH
VHDCDSBAGLKGNAERHEIRLPAKATVAEPAPVHEDYAGEASVRHETKVSANDNF
LLKHQTDLIKNHNMNTPPKUSFEITKSEITKLISDGIARAALAEESLLRNPNLNPN
HPNQNPINPHQKNVPTTEISSRPEILLGTHAHTLNASPSAYSPPSATLPSANL
KLPIAAGPHLTIHHPHLHRLHQPHAPHPVHLHHNLTAFLPALAQKTIGLSEFL
RPPONAPLFPVKLPGRPFAPKIKVPASRPILPQHPHLHPHPOOHVPLLOQOQPSL
IVSHYKPIPGLLKPFVKEKPEPQPLAASVLLGQTELGGLNKNKGRKLTGKPKI
PVPYVDLEPQSLQNTAIFNPGGKGGKQDKPKASSVSISTPTPLVXRPVTKPSQE
ETASMPAVNQGFPDPTVIVESGFKPIVTRDTGTGVLQPKETIIDQVAHREDPDGTEIDE
VMTDTLFLAAQOQSGSESPFEPMPISPLDSTNATKVLRVNKEVPSATSLRPSA
ALEHALPSASELIKPTLDELFAEELNEELEMEMPVADVDESLETTKKDAVTTIKIP
IPRNTTKKPDOLLDELFGDDEELYADELEMDRDDRMAAAARIDTVYLPDNRKIP
DTPVPSGALYTFDCKSVSDSILVPPKLPDAPDNANVHORHAQYGLTPLEQLVRRTPQF
GYVGLPQERFCTEPQVPSYSHIPKPSRTPVPFSSSGSTIYPYSSSTGASTSVS
SSASPLSSSLRPISTKLQLLKPEGRRA"
join(51689. .51886,54866. .54906,54969. .55179,55303. .55774,
56086. .57923,57990. .>58232)
/gene="pug"
/product="CT13530"
/db_xref="FLYBASE:FBan0004067"
<51689. .>58232
/gene="pug"
/note="CG4067"
/map="86C2-86C2"
/db_xref="FLYBASE:FBan0004067"
/db_xref="FBgn0020385"
join(51782. .51886,52710. .>54028)

Alignment Scores:
Pred. No.: 11.4 Length: 224795
Score: 90.00 Matches: 30
Percent Similarity: 45.10% Conservative: 16
Best Local Similarity: 29.41% Mismatches: 56
Query Match: 11.58% Indels: 0
DB: 3 Gaps: 0

US-09-847-539a-6 (1-159) x AE003688 (1-224795)
QY 26 GluLysLeuAlaLeuArgAsnGluGluAtrgAlaIleAspGluLeuLysGlnAlaIle 45
Db 124080 GAAAGCAGCGTCTGCTAAGCCAAAAGCGTCGCGAGGAGGCCCAAAATCGCTGCC 124021
QY 46 GluAspLysGluAlaThrThrAlaIleGluAlaAlaSerSerAspAlaLeuGluAlaLeu 65
Db 124020 GAAAGGCTCTCCTCGCTGAAGCTGCCGCCCAAAAAGCAGCGAGCTAAGCCCTTA 123961
QY 66 AlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValIysAlaAspAsnAla 85
Db 123960 AAAGCGCTGAAGATGCTCCCGAAGAGCGGCTGCGGAAGAGCGGCTACGGAGGAGGCT 123901
QY 86 AlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAla 105
Db 123900 GCCGCCCAAAAGTCCCGAGGAAGCTGCCCAAAAAGCTGCTGAGGAAGCTCGCTAGCA 123841
```



Kurokawa, K., Yasunaga, T., Yokoyama, K., Makino, K., Shinagawa, H. and Hayashi, T.  
Comparative analysis of the whole set of rRNA operons between an enterohemorrhagic *Escherichia coli* O157:H7 Sakai strain and an *Escherichia coli* K-12 strain MGL1655  
Syst. Appl. Microbiol. 23 (3), 315-324 (2000)  
20557356

3 (sites)  
Yokoyama, K., Makino, K., Kubota, Y., Watanabe, M., Kimura, S., Yutsudo, C. H., Kurokawa, K., Ishii, K., Hattori, M., Abe, H., Iida, T., Yamamoto, K., Hayashi, T., Yasunaga, T., Honda, T., Sasakawa, C. and Shinagawa, H.  
Complete nucleotide sequence of the prophage  $\phi$ 1-Sakai carrying the Shiga toxin 1 genes of the enterohemorrhagic *Escherichia coli* O157:H7 strain derived from the Sakai outbreak  
Gene 258 (1-2), 127-139 (2000)  
20564182

4 (sites)  
Hayashi, T., Makino, K., Ohnishi, M., Kurokawa, K., Ishii, K., Yokoyama, K., Han, C.-G., Ohtsubo, E., Nakayama, K., Murata, T., Tanaka, M., Tobe, T., Iida, T., Takami, H., Honda, T., Sasakawa, C., Ogasawara, N., Yasunaga, T., Kuhara, S., Shiba, T., Hattori, M. and Shinagawa, H.  
Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12  
DNA Res. 8 (1), 11-22 (2001)  
21156231

5 (bases 1 to 222605)  
Ohnishi, M., Kurokawa, K., Makino, K., Yasunaga, T., Shinagawa, H. and Hayashi, T.  
Direct Submission  
Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome Information Research Center; 3-1, Yamadaoka, Suita, Osaka 565-0871, Japan (E-mail: ken@gen-info.osaka-u.ac.jp, URL: http://www.gen-info.osaka-u.ac.jp/, Tel: 81-6-6879-8365, Fax: 81-6-6879-2047)  
genome project.

Location/Qualifiers  
1..222605  
/organism="Escherichia coli O157:H7"  
/strain="O157:H7"  
/sub\_strain="RIMD 0509952"  
/db\_xref="taxon:83334"  
74..607  
/gene="ECs1423"  
74..607  
/gene="ECs1423"  
/note="similar to B1045\_ECOLI g11787283 percent identity 100 in 177 aa (Conserved in E.coli K-12)"  
/codon\_start=1  
/transl\_table=11  
/evidence=not\_experimental  
/product="putative polyprotein"  
/protein\_id="BAB34846.1"  
/db\_xref="GI:13360887"  
/translation="MKTRHVVQGDITKLAVDIVYNAANPSLMGGGVGDGATHRAAGP ALLDLKVRQGGDCPTGHAVITLAGDLPKAVVHTVGPVWRGGEQEDQLLDQAYL NSRLVANSYTSVAFPALSTGVYGPRAAAAEIAVKTVSEFTRHALPEQVYFYCYD ENAHLYERLLTQQGDE"  
549..2030  
/gene="ECs1424"  
549..2030  
/gene="ECs1424"  
/note="similar to YMDC\_ECOLI g11787284 percent identity 99 in 493 aa (Conserved in E.coli K-12)"  
/codon\_start=1  
/transl\_table=11  
/evidence=not\_experimental  
/product="putative synthase"  
/protein\_id="BAB34847.1"  
/db\_xref="GI:13360888"  
/translation="MMKKTPTSTKDSLPMKEMNDLPRLASVPLCSQHPGCGGLFPL EKSLDAFAARYLAEMSEHTLDVQYIWIQDDMSGRLLFSALLAAAKRGVVRVRLLLDDN

NTPGDLDIRLLDSDHPRIEVRLEPNPFSRLLRLPLGYITDFSRLLRRMHNKSFTVDGVV  
TVGGRNIGDAYGEAEPLFSDLDVMAIPVPEVDVADDFARYWYCKSVPLSQDLDV  
PTEGMADRIELPASWNHNDAMTHRYLRKMESSPFIHLVDGTPLTWAKVRLSSDDPAK  
GEGKAKRSLLPQRLFDINGSPSERIDIISYFVPTRAGVAQLLBMVKGKVIATLTN  
SLAANDVAVHAGYARWRKKLRJYVELVEYETSETLDRGDTGNSGASLIANKT  
FSIDGKTIVGIFNDFPSTLLNTMGFVIESETLAQLDKRFIOSQYDAAWOLRJDJR  
WGRINNVDRHAKKEIVLKKEPATSFMKRVNVRASILPVEWLL"  
complement(2038..3195)  
/gene="ECs1425"  
complement(2038..3195)  
/gene="ECs1425"  
/note="similar to MDOC\_ECOLI g113025226sp1p75920 percent  
identity 99 in 385 aa (Conserved in E.coli K-12)"  
/codon\_start=1  
/transl\_table=11  
/evidence=not\_experimental  
/product="glucans biosynthesis protein"  
/protein\_id="BAB34848.1"  
/db\_xref="GI:13360889"  
/translation="MNPVPAQREYFLDSTRAMLLGIPFHSILVSSHTWVNSAEP  
SLMLTLFNDFIHSFQMOVEFVLSGVESYMLFLRYPLKWKVYRVERVGIPLMTAIPLL  
TUPQIMLOYVKGKAEWPGSLIDYKNTLAWELISHLWFLVLVYVMTLCLVWIFKRI  
RNNLESDQYVAGKASWPKLSVIFLCIGIGYAVIRRTIFIVYPIILNSGTFNFIVMOTL  
FYLPFFILGALAFIPFHLKALFTTPSRGCTLAALAFAYLLNQRYSGSDAMMYETES  
VITMVLGLMWNVVFSGHRLNFQSARVTVFNASLFYLVVHHPJTLFCFGAYITPHI  
TSNMLGFLCGLIFVVGIAIILYELHILRIPLLAFSGKPKVVKRENDKAPAR"  
3588..5123  
/gene="ECs1426"  
3588..5123  
/gene="ECs1426"  
/note="similar to MDOG\_ECOLI g11787286 percent identity  
99 in 511 aa (Conserved in E.coli K-12)"  
/codon\_start=1  
/transl\_table=11  
/evidence=not\_experimental  
/product="periplasmic glucans biosynthesis protein"  
/protein\_id="BAB34849.1"  
/db\_xref="GI:13360890"  
/translation="MMKMRWLSAAVWLTLTYTSSNAFSDIDVAKAQAQSLAGKGYEAPK  
SNLPSVFRDMKYADYQIQFNHDKRAYNNLTKYFKLEFYHOGMYEDTPVKINEVATA  
VKRIKYPDYFTFGDVQHDKDTVKDLGAGFKVLYPINSKDKNDIVSLMGASYFRVI  
GAGQVYGLSARGLAIDTALPSGEEFPRKPEWIERPKDKRLTIVALLDSDPRATGAY  
KEVHAPGRDVTVDVQSKIYLRDKVGKLVGAPLITSMFLPGNPQSPANNYRPELHDSNG  
LSIHAGNCEIMWRPNKPLAVSFSMENPOGFLJORGDRFSRPEDLDYDLRPS  
AWYTPKGEWKGKSVELVEIPTNDEINDNIVATWTDLPQPEKEMNFKYITIFSRDED  
KLHAPDNAMVQQTTRSTGDVKOSNIROPDQIAFVVDFTGAEMKRLPDDTPVTQTS  
IGDNGEIVESTVRYNPVTGKRWLRVVKVDAKKTTEMRAALVNADQTLSETWSYQLP  
ANE"  
5086..7659  
/gene="ECs1427"  
5086..7659  
/gene="ECs1427"  
/note="similar to membrane glycosyltransferase MDOH\_ECOLI  
g11787287 percent identity 100 in 847 aa but differs at  
N-ter (Conserved in E.coli K-12)"  
/codon\_start=1  
/transl\_table=11  
/evidence=not\_experimental  
/product="membrane glycosyltransferase"  
/protein\_id="BAB34850.1"  
/db\_xref="GI:13360891"  
/translation="MKPGATSYLPMNKTTEYIDAMPIAASEKAALPKTDIRAVHONLD  
ADHRTWAKEDDPSQGSVKARLEQWPSLADQGLIKDDGRQLKAMPKAKSSMPFD  
PHRTNPVGRFNDRLGRDVTPLYARLTKEQESQKRWRTVGTIRYIILITLAQTV  
VATWTKTILPYQGWALLNPMQMDVQDLWSPMQLLYMLOTGILILFAVLFCWVSAG  
FWTALMGFLQLLIGRDYKISASATVGEPLNPHRTALIMPTCNEDVNRVFAGLRATW  
ESVKATGNAKHFDVYILSDYNPDICVAEQKAWMLIAEVGEGGOIFYRRRRVRKRK  
SCNIDDCRRWCSQSYWVYLDADSVYMGDCLGLVRLMEANPNAGIIOSSPKASGMD  
TLYARQQQFATRYGPTAGHFWOLGESHYSHYNATIRVKKFTEHCALAPLPGEGS  
FAGSLSHDFVEAALMRKAGWGIAYDLPGSYBELPNLDELDRMCHGNLWNF  
RFLVKGHPHVAFLVGTVMVSYLSAPLWFMFLALSTALQVVAHALTEPOYFQPLQFL  
PWVQWRLELAIALFASTMVLLFLPKLLSILLIWCCKGTKEYGGFVRTVLSLLEVLFS  
VLLAPVRMLFHTVVFVYSAFLGWVYVWNSPORDDDSTSGCEAFKRGHSQILLICLVWAVG



RESULT	29
AC017732/c	
LOCUS	54784 bp DNA linear HTG 10-DEC-1999
DEFINITION	Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered pieces.
ACCESSION	AC017732
VERSION	AC017732.1 GI:6554267
KEYWORDS	HTG; HTGS_PHASE2.
SOURCE	fruit fly.
ORGANISM	Drosophila melanogaster
REFERENCE	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
AUTHORS	1 (bases 1 to 54784)
TITLE	Adams,M. and Venter,J.C.
JOURNAL	Direct Submission
COMMENT	Submitted (10-DEC-1999) Celera Genomics, 45 West Gude Drive, Rockville, MD, USA This sequence was identified as CDM:10211903 by the submitter. For more information on this record e-mail to fly@celera.com. * NOTE: This is a 'working draft' sequence. * This sequence will be replaced * by the finished sequence as soon as it is available and * the accession number will be preserved.
FEATURES	Location/Qualifiers
source	1..54784
/organism="Drosophila melanogaster"	
/db_xref="taxon:7227"	
BASE COUNT	15347 a 11327 c 11357 g 16753 t
ORIGIN	
Alignment Scores:	
Pred. No.:	2.48 Length: 54784
Score:	90.00 Matches: 30
Percent Similarity:	45.10% Conservative: 16
Best Local Similarity:	29.41% Mismatches: 56
Query Match:	11.58% Indels: 0
DB:	Gaps: 0
US-09-847-539A-6 (1-159) x AC017732 (1-54784)	
QY 26 GluLysLeuAlaLeuArgAsnGluGluArgAlaIleAspGluLeuLysGlnAlaIle 45	
Db 20122 GAAAGCAGCGTCTGCTGAAGGCCAACAAAGGCTCCGAGGAGCGCGCAATTCGTGC 20063	:    :    :    :    :
QY 46 GluAspLysGluAlaThrThrAlaIleGluAlaSerSerAspAlaLeuGluAlaLeu 65	:    :    :    :    :    :
Db 20062 GAGAGGCTCTCTGGCTGAAGCTGGCGCCCAAAGACGCCGAAGAAGCTTAAGCCCTA 20003	:    :    :    :    :    :
QY 66 AlaAspGlnThrAspAlaLeuGlnSerGluGluAlaValVallLysAlaAspAsnAla 85	:    :    :    :
Db 20002 AAAGCCGCTCAAGATGCTGCCAGAAAGCTGCCGAAGACCGCTTAGCGGAGGAGGCT 19943	:    :    :    :    :    :
QY 86 AlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAla 105	:    :    :    :    :    :
Db 19942 GC GGCCCCAAAAGGTGCGCGAGGAAGCTGCCAAAAAGCTGCTGAGGAAGCTCGCCTAGCA 19883	:    :    :    :    :    :
QY 106 GluValValGlnSerAspAsnAlaAlaSerAspAlaTrpGluLysAlaAlaThrProile 125	:    :    :    :    :    :
Db 19882 GAGGAGCGACCGCCCAAGAGCGCGGAGGAAGCTGCCCAAAAGGCTGCCGAAGAAGCC 19823	:    :    :    :    :    :
QY 126 AlaLeu 127	
Db 19822 GCTTTA 19817	
RESULT 30	
AC008140/c	
LOCUS	140973 bp DNA linear INV 24-FEB-2001
DEFINITION	Drosophila melanogaster, chromosome 3R, region 86C-86C, BAC clone BACR15J23, complete sequence.
ACCESSION	AC008140
VERSION	AC008140.10 GI:13127273
KEYWORDS	HTG.

```

SLRLTSGKGTDLARLVTTAIDTAIDTEHFRSGLPLGEYTLTVRAINSYGOQGEPAITTFRI
AAPAPSRILETFCYQITATPLHAVIDPTQVQEFWFSEKRIADIQVETAARYLGS
LWYIAASINLPGHDIYFYIRSVNTVKGSAFVAAGVRSADAGCYLDFKFGKIGKTHL
AQLWTQIDNQLQADPLAEIRTSITVNSNETITQVKNKLENOQSAALQOIQKQVODTNN
NLMSMAVKLQOQDGLXYTAGIGAGIENTPAGCQSLQALLAADRIAMINPANGTTPM
FVQGOQOIFMNEVFLXYLTAPTITSGNPPAFSLTDPGLTAKNADISGNVANSSTGL
NNVTINENCVRGLKLANQLEGDLVKTGKAFPRDSRAPERWPSGTTTIVRYVDQDPF
RQVITPAVASGAKHEKHTDIYSSCLIVRKNGAEIYNRTALDNTLIYSGVIDMPAG
HGHTLEFSYSAWLNVNWPYASISIDLVVVMKATAGITIS"
/ gene 3661. .4260
/ gene="21917"
/ CDS 3661. .4260
/ gene="21917"
/ function="putative membrane; Other or unknown (Phage or
Prophage Related)"
/ note="Residues 1 to 199 of 199 are 75.87 pct identical to
residues 1 to 199 of 199 from GenPept 118 :
g117532789|gb|AA63231.1|AF151091_2 (AF151091) Lom
[prophage P-E1bA]"
/ codon_start=1
/ transl_table=11
/ product="putative outer membrane protein of prophage
CP-933X"
/ protein_id="AAG56006.1"
/ db_xref="GI:12514846"
/ translation="MRKLCVILSAVVMQVAAATPASAAHQSTLSAGYLHASTNVPAG
SDLLGINVRYEFMDALGLITSEFVNAEDQKTRYSDTRWHEDSVRRWRFVSVAG
PSVRYNEFSAYAMAGVAYSRSVTFESGDYLRVDNKGKTHDVLTSDDGRHSWTSLAW
GAGVQNPXESVALIDIAIEGSGSGDWRTDGFIVGVGYK"
/ gene 4319. .7240
/ gene="21918"
/ CDS 4319. .7240
/ gene="21918"
/ function="putative membrane; Not classified (Phage or
Prophage Related)"
/ note="Residues 1 to 498 of 973 are 95.78 pct identical to
residues 1 to 498 of 1122 from GenPept 118 :
g117873636|gb|AAC74454.1| (AE000234) putative membrane
protein [Escherichia coli]"
/ codon_start=1
/ transl_table=11
/ product="putative membrane protein of prophage CP-933X"
/ protein_id="AAG56007.1"
/ db_xref="GI:12514847"
/ translation="MNAVIGSVLKDGTGKPVENCTIOLKARRNSATVVVNTVASEN
PDEAGRYSMDEVYGOYSVILLVEGFPPSHAGTITVYEDSOPGTLNDFLGAMTEDDVRP
EALRFLMVEVARNASAYANTAAAKKSASDASARSAREATHADAARSARASTS
AGQAASQASASSAGTASTKATEASKSAASAAESSKSAATSAGAAKTSETNAASVQO
SAITSASTATTKASEAASSASDASAKSAESKSETSAASSASASATAAGNSAKAA
KTSSTNAKSETAEQASAAAGSKTAAALSANAATSAGQASASATAAGKSAESAA
SASTATTKAGATEQASAAASASAAKTSETNAKASSETSSESKTAAASASSAAGSA
SSASAKDEATROASAAKSSATTASTKATEAGSATAAAOQSTAESATRAETAAR
AEDIASVALEADSTTKGIVOLSSATNSESLSAATPKAVKAAAYELANKGYTAQDAT
TAQKIVOLSNATNSTEMLAATPKSVKAAAYDLANGITADQATTAQKIVOLSSATN
SASTLAATPKAVKAAANDNANGRPVSARKVNGKALSSDITLTPKDTGLNSTNFSG
GAGFKLATVTPQASSVNSITLIGAGFNVPQOQAGISLVLIRAGNPKGITGAL
WORTSTGTFNFWNTSGDYDIYVAGNATVNIOWDYTSNASVTIHTSPVYANK
PEGITDCTVSLYTPSPQFPYPCGAPIPWSDTVPSPGYALMGOTFQKSAVPKLAAYP
SGVTPDMRGHTIKGKPSAGRAVLSEQEDGIKSHTHSASASSTDLTKTTSFDPYTKS
TNNTGATHSVSGTAAGNHTHSVTGASAVSQWSQNGSVHKVVSAASTVNSRAGHT
HSVSGTAAASGAHAHTVCGIAGHTHSVAIGSHGHTITVNAAGNAENTVKNIAFYIVRL
A"
/ complement(4822. .5571)
/ gene="21919"
/ CDS complement(4822. .5571)
/ gene="21919"
/ function="orf; Other or unknown (Phage or Prophage
Related)"
/ note="Residues 45 to 248 of 249 are 64.21 pct identical
to residues 5 to 205 of 206 from GenPept 118 :
g115089911|gb|AA96556.1| (J02459) orf206b [bacteriophage
lambda]"
/ codon_start=1

```

```

/transl_table=11
/product="unknown protein encoded by prophage CP-933X"
/ protein_id="AAG56008.1"
/ db_xref="GI:12514848"
/ translation="MDAVVALLFAADACLVASSFEADADDADAELADDAADAAVFEDS
ALVSDPAFVSDVFAAEAEIAALACSVASPAFVVAVEADDAALSDNDFPAVALAEAC
PALVDAALADANAAPFEPAPAEALCSAAVSEDLAFVSDVLAFAELPAAVBEAALLA
LEAALVSDDFAPASFEADASLAENASDAFVAVDAEVAADCCDTAAFAVSDVFAAPALV
AAALFSDSAAAALFDASVAFVDAVPALEAD"
/ gene 7240. .7821
/ gene="21920"
/ CDS 7240. .7821
/ gene="21920"
/ function="orf; Other or unknown (Phage or Prophage
Related)"
/ note="Residues 1 to 191 of 193 are 94.76 pct identical to
residues 1 to 191 of 191 from GenPept 118 :
g112367120|gb|AAC74619.1| (AE000252) orf, hypothetical
protein [Escherichia coli]"
/ codon_start=1
/ transl_table=11
/ product="putative tail fiber protein of prophage CP-933X"
/ protein_id="AAG56009.1"
/ db_xref="GI:12514849"
/ translation="MAFRMSEOPTRTIKYINLLTGTNEFIGEGDAYIPPHGTGLPANSTY
IAPDPIAGFVAVVNSDESKWHLVEDHREKTVYDVASGDALFISELGSPLPNVTWLS
EGEFQKWNCTAVVKDAEAKLFRIREAETKNSLMQVASEHIAPIQDAVDLEIATEEE
ASLLEAKKRYRLNVRDSTAPDIEMPIEV"
/ complement(7941. .8831)
/ gene 7941. .8831
/ CDS complement(7941. .8831)
/ gene="21921"
/ function="orf; Other or unknown (Phage or Prophage
Related)"
/ note="Residues 1 to 289 of 296 are 84.42 pct identical to
residues 1 to 289 of 292 from GenPept 118 :
g117162108|emb|CAB76676.1| (Y19180) putative catalase
[Salmonella typhimurium]"
/ codon_start=1
/ transl_table=11
/ product="unknown protein encoded by prophage CP-933X"
/ protein_id="AAG56010.1"

```

## Alignment Scores:

Pred. No.:	0.416	Length:	10432
Score:	90.00	Matches:	29
Percent Similarity:	45.92%	Conservative:	16
Best Local Similarity:	29.59%	Mismatches:	52
Query Match:	11.58%	Indels:	1
DB:	1	Gaps:	0

US-09-847-539A-6 (1-159) x AE005333 (1-10432)

Qy	50	AlaThrThrAlaLeuGluAlaAaSerSerAspAlaLeuGluAlaLeuAlaAaSpGlnThr	69
Db	5375	GCAGCGAGTTCTGTCGCCGAGCAAGACGCCGAAACGCGTCCGAAACC	5434
Qy	70	AspAlaLeuGlnSerGluAlaAlaValLysAlaAspAsnAlaAaSerAspAla	89
Db	5435	AGCCGAGAATCCTCAAAACGGCTGCGGCATCGTCAGCCAGTTCGCGCGGCTCATCGCA	5494
Qy	90	LeuGluAlaLeuAlaAaSpGlnThrAspAlaLeuGlnSerGluAlaAlaValValGln	109
Db	5495	TCATCTCGCTGCTGTCCTCAAGAGATGAGCGACAGCAAGCGTCAGCAGCGAGCAGC	5554
Qy	110	SerAspAsnAlaAaSerAspAlaTrpGluLysAlaAlaThrProIleAlaLeuAspVal	129
Db	5555	GCCACGACGGCATCCACGAGGCGACAGCGAGCTGGTAGTCGACGCGAGCTCAG	5614
Qy	130	LysLysThrLysAspThrLysProValLysLysGlu-GluArgGlnAsn	146
Db	5615	AGCAAAAGTACGCGGGGAATCTGCAGCAACGCGCGCTGAGACAGCGCGCAAAAC	5666







```
|||||
Db 1416 GCAACTCTCTACAACTGGTGAAGGAACCAACCA 1451
|||||
RESULT 24
SG148IGG
LOCUS
DEFINITION SG148IGG 1469 bp DNA linear BCT 12-JUL-1995
ACCESSION X04015
VERSION X04015.1 GI:47071
KEYWORDS gamma-immunoglobulin binding protein G.
SOURCE Streptococcus sp.
ORGANISM Streptococcus.
REFERENCE 1 (bases 1 to 1469)
AUTHORS Guss,B., Eliasson,M., Olsson,A., Uhlen,M., Frej,A.K., Jorvall,H.,
Flock,J.I. and Lindberg,M.
TITLE Structure of the IgG-binding regions of streptococcal protein G
MEDLINE EMBO J. 5 (7), 1567-1575 (1986)
COMMENT Repetitive structures of 24 amino acids, A1, A2 and A3, are found
interrupted by two unrelated repeats of 51 amino acids, B1 and B2.
Further downstream, following a 12-residue unique "spacer" region'
(S), there is a stretch of 55 amino acids repeated three times, C1,
C2 and C3. The latter regions are interrupted by two 15-residue
regions, called D1 and D2. Directly following these repeats there
is a region W, which consists of a repetitive, extremely
hydrophilic, structure containing several proline residues. This
region resembles a similar structure in the streptococcal M protein
and in protein A, which is thought to mediate the binding of the
protein to the cell wall.
FEATURES
source Location/Qualifiers
CDS
1..1469
/organism="Streptococcus sp."
/strain="G 148"
/db_xref="taxon:1306"
<1..1443
/codon_start=1
/transl_table=11
/product="protein G"
/protein_id="CAA27638.1"
/db_xref="GI:47072"
/db_xref="SWISS-PROT:PI1909"
/translation="EFNKYGVSDDYKLNINNAKTVEGKDLQAOVVESAKKARISEAT
DGLSDFLKSGTQPAEDTVKSIELAEAKVLANRELDKYGVSDDY
QAOVVESAKKARISEATDGLSDFLKSGTQPAEDTVKSIELAEAKVLANRELDKYGVSDDY
YKLNINNAKTVEGKALIDEILAAALPKTDYTKLILNGKTLKGETTTEAVDAATAEKVF
KOYANDNGVGEWYDDATKFTVTEKPEVIDASELTPAVTYTKLV
EAVDAATAEKVFQKQYANDNGVGEWYDDATKFTVTEKPEVIDASELTPAVTYTKLV
INGKTLKGETTTEAVDAATAEKVFQKQYANDNGVGEWYDDATKFTVTEKPEVIDASELTPAVTYTKLV
APTEPEKPEASIPVLPTPATPIAKDDAKKDDTKKEDAKKPEAKKEDAKKAEETLPTTG
EGSNPFETAAALAVMAGALAVASRRKED"
BASE COUNT 546 a 279 c 297 g 347 t
ORIGIN
Alignment Scores:
Pred. No.: 0.0504 Length: 1469
Score: 90.00 Matches: 18
Percent Similarity: 62.50% Conservative: 2
Best Local Similarity: 56.25% Mismatches: 12
Query Match: 11.58% Indels: 0
DB: 1 Gaps: 0
US-09-847-539a-6 (1-159) x SG148IGG (1-1469)
Qy 128 AspValLysLysThrLysAspThrLysProValLysLysGluGluArgGlnAsnVal 147
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1270 GATACTAAGAAAGAGATGCTAAACCAAGAGCTAAGAAAGAGCTAAGAAAGCT 1329
Qy 148 AsnThrLeuProThrThrGlyGluGluSerAsnPro 159
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1330 GAAACTCTCTACAACTGGTGAAGGAACCAACCA 1365
Alignment Scores:
Pred. No.: 0.0504 Length: 1469
Score: 90.00 Matches: 18
Percent Similarity: 62.50% Conservative: 2
Best Local Similarity: 56.25% Mismatches: 12
Query Match: 11.58% Indels: 0
DB: 1 Gaps: 0
US-09-847-539a-6 (1-159) x SG148IGG (1-1469)
Qy 128 AspValLysLysThrLysAspThrLysProValLysLysGluGluArgGlnAsnVal 147
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1270 GATACTAAGAAAGAGATGCTAAACCAAGAGCTAAGAAAGAGCTAAGAAAGCT 1329
Qy 148 AsnThrLeuProThrThrGlyGluGluSerAsnPro 159
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1330 GAAACTCTCTACAACTGGTGAAGGAACCAACCA 1365
|||||
Db 1416 GCAACTCTCTACAACTGGTGAAGGAACCAACCA 1451
|||||
RESULT 25
AL2446
LOCUS
DEFINITION AL2446 1469 bp DNA linear PAT 04-JAN-1994
ACCESSION AL2446
VERSION AL2446.1 GI:491352
KEYWORDS synthetic construct.
SOURCE synthetic construct.
ORGANISM synthetic construct.
REFERENCE 1 (bases 1 to 1469)
AUTHORS
TITLE METHOD AND MEANS FOR PRODUCING A PROTEIN HAVING THE SAME IgG
SPECIFICITY AS PROTEIN G
PATENT: WO 8705631-A 4 24-SEP-1987;
FEATURES
source Location/Qualifiers
CDS
1..1443
/organism="synthetic construct"
/db_xref="taxon:32630"
1..1443
/notes="Protein sequence is in conflict with the conceptual
translation"
/codon_start=1
/transl_table=11
/product="Protein having the same IgG specificity as
protein G from group G streptococci strain G148"
/protein_id="CAA01034.1"
/db_xref="GI:4526638"
/translation="EFNKYGVSDDYKLNINNAKTVEGKDLQAOVVESAKKARISEAT
DGLSDFLKSGTQPAEDTVKSIELAEAKVLANRELDKYGVSDDYHKNLNINNAKTVEGKDL
QAOVVESAKKARISEATDGLSDFLKSGTQPAEDTVKSIELAEAKVLANRELDKYGVSDDY
YKLNINNAKTVEGKALIDEILAAALPKTDYTKLILNGKTLKGETTTEAVDAATAEKVF
KOYANDNGVGEWYDDATKFTVTEKPEVIDASELTPAVTYTKLVINGKTLKGETT
EAVDAATAEKVFQKQYANDNGVGEWYDDATKFTVTEKPEVIDASELTPAVTYTKLV
INGKTLKGETTTEAVDAATAEKVFQKQYANDNGVGEWYDDATKFTVTEKPEVIDASELTPAVTYTKLV
APTEPEKPEASIPVLPTPATPIAKDDAKKDDTKKEDAKKPEAKKEDAKKAEETLPTTG
EGSNPFETAAALAVMAGALAVASRRKED"
BASE COUNT 545 a 280 c 297 g 347 t
ORIGIN
Alignment Scores:
Pred. No.: 0.0504 Length: 1469
Score: 90.00 Matches: 18
Percent Similarity: 62.50% Conservative: 2
Best Local Similarity: 56.25% Mismatches: 12
Query Match: 11.58% Indels: 0
DB: 6 Gaps: 0
US-09-847-539a-6 (1-159) x AL2446 (1-1469)
Qy 128 AspValLysLysThrLysAspThrLysProValLysLysGluGluArgGlnAsnVal 147
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1270 GATACTAAGAAAGAGATGCTAAACCAAGAGCTAAGAAAGAGCTAAGAAAGCT 1329
Qy 148 AsnThrLeuProThrThrGlyGluGluSerAsnPro 159
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1330 GAAACTCTCTACAACTGGTGAAGGAACCAACCA 1365
Alignment Scores:
Pred. No.: 0.0504 Length: 1469
Score: 90.00 Matches: 18
Percent Similarity: 62.50% Conservative: 2
Best Local Similarity: 56.25% Mismatches: 12
Query Match: 11.58% Indels: 0
DB: 1 Gaps: 0
US-09-847-539a-6 (1-159) x SG148IGG (1-1469)
Qy 128 AspValLysLysThrLysAspThrLysProValLysLysGluGluArgGlnAsnVal 147
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1270 GATACTAAGAAAGAGATGCTAAACCAAGAGCTAAGAAAGAGCTAAGAAAGCT 1329
Qy 148 AsnThrLeuProThrThrGlyGluGluSerAsnPro 159
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1330 GAAACTCTCTACAACTGGTGAAGGAACCAACCA 1365
|||||
LOCUS
DEFINITION AY060997 2022 bp mRNA linear INV 08-NOV-2001
ACCESSION AY060997
VERSION AY060997.1 GI:16768651
KEYWORDS fruit fly.
SOURCE Drosophila melanogaster
ORGANISM Drosophila melanogaster
REFERENCE 1 (bases 1 to 2022)
AUTHORS Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
Stapleton,M., Brokstein,P., Hong,L., Agbayani,A., Carlson,J.,
Champe,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E., George,R.,
```



```

/db_xref="GI:13360214"
/translation="MRGKIPKTELLVTFEVARHESYTRAAEELALTQSAFVROVSAL
EERLTPLESHSKRIFLNDTKTYLGIYKVELTKLERDTNTIMTQPTQVQVLELAVN
PIFSEHNLPLNHEFTKLPDIIIVNIHSLANGDFLNRREYDAVIRENFCAPAEVEY
LPEEFELPVCSSLLAMSQKSVAEELLFELLHQSSTRITGWEEFWFALSGVSSPLVN
NGPRFDLSMLIAAARNSIGNVALLPRFAIQTDLSDGMDVPCDVPRTGNRRITMTQEE
EKSDSPHLQGFREMLLAKSVVPEM"
complement(3720..4322)
/gene="ECs0756"
complement(3720..4322)
/gene="ECs0756"
/notes="probable cob(I)alamin adenosyltransferase, similar
to cob(I)alamin adenosyltransferases (corrinoid
adenosyltransferases) e.g. [Escherichia coli]
gi1115148|sp|P13040|BTUR_ECOLI percent identity 67 in 200
aa"
/codon_start=1
/transl_table=11
/evidence="not_experimental"
/product="putative cob(I)alamin adenosyltransferase"
/protein_id="BAB34179.1"
/db_xref="GI:13360215"
/translation="MEARISTEHRVROOKLEQVDVTRVAAAEIKKGILIVTGNKGK
KSTAFGTTRAVGKTKTVGAQYIKGWDNGEYNLLQPLGVFEFHMGFTGTWETQNR
QADIDAAKEVSWESKRMADKRYDLVLDLTYMLAYHYLDTEEVIASLQNRPAQQSV
IVTGRCHSOLIKMADTVSEIRPVKHAFDNGIOAQPGIDW"
complement(4332..5984)
/gene="ECs0757"
complement(4332..5984)
/gene="ECs0757"
/notes="probable fumarate hydratase, similar to fumarate
hydratases e.g. fumarate hydratase class I, aerobic
(fumarase) - Escherichia coli
gi1120598|sp|P0923|FUMA_ECOLI percent identity 68 in 545
aa"
/codon_start=1
/transl_table=11
/evidence="not_experimental"
/product="putative fumarate hydratase"
/protein_id="BAB34180.1"
/db_xref="GI:13360216"
/translation="MSKPIFQWPFLOKNDGTEYTLISDQHTVTETDGEVYKIAPE
ALTLLARQAFYAFPLFSARHQAQVAILDQOASNDKYVALQLLRNAEVSAGKVP
NQCDTQATIVASKGOQIWTCGNDAAEALSGIYSTFOENLRFSONALPDMTEVNTQ
TNLPAQIDISAVAGDEYHFLCVNKGGSANKALQETKSLQPEKLTAFLEKKMSL
GTAACPPIHIAVVGSLSDAQDLTKIAKASTYTDNLPTSGNEQGOAFRDIKELKVL
EASQFGIGAGGGKYFAHDIRIVLRHGGSCPTAMALSCADRNIKAKINKHGLW
EXLEHNPQYIPASLREENHAQHVQDLNRLPRDVMQDLARLPVGRVLSQPIVVAR
DIAHAKIKARLDSGEMPEYLLKHHIVVYAGPAKTPENMACGSLGPTTGGRMGGYIDTF
QAAGSLVMSLXGNBSQOVTDACHKHGFNLGICGAALLAQAEYVKSRLCLYPELPG
MEAVNMHEVNPAPFILVDDKGNNEFSQFEQQHRCASCAPGH"
complement(6092..7372)
/gene="ECs0758"
complement(6092..7372)
/gene="ECs0758"
/notes="probable transport protein, similar to
glutamate/aspartate transport proteins (proton glutamate
symport proteins) e.g. [Bacillus stearothermophilus]
gi1121467|sp|P24943|GLTT_BACST percent identity 38 in 416
aa, also similar to C4-dicarboxylate transpor"
/codon_start=1
/transl_table=11
/evidence="not_experimental"
/product="putative transport protein"
/protein_id="BAB34181.1"
/db_xref="GI:13360217"
/translation="MKKISLTMTMILLALVLMIGVLNNTASPETAKLYAOEISFT
TIFLRKMLIAPLVYTLVGIKAWGDAKALGRIFSKFLFICALLSIALGLITY
NFEFMGTGINFVAHAGETGVVADEFTLKFVISHAFPTSIDMAHNEILQIVFSI
FVGSCLTAIGKSAITVHALDSLAMKLTGYMLFAPLTFVFAISALIAERGLAVM
VAGITPMGSEFYTMLLWLLGLAIYVYVPCIRTRALSEPALLAFTTSSEAAFP
GTLEKLEQFGSPKVPICVYGFNLVSGMAYGSFATVFAQACNIHLSLSTGEQITM
LLILMLTSKMGAGVPRASHVIAATLNQNIPEAGILLMGVDPFLDMGRSRTNWSN

```

```

gene
AMCAAMVSRWEGEHFGEGRGKALKPNESNALP"
complement(7533..7850)
/gene="ECs0759"
CDS
complement(7533..7850)
/gene="ECs0759"
/notes="unknown"
/codon_start=1
/transl_table=11
/evidence="not_experimental"
/product="hypothetical protein"
/protein_id="BAB34182.1"

Alignment Scores:
Pred. No.: 2.6 Length: 297816
Score: 96.00 Matches: 27
Percent Similarity: 51.47% Conservative: 8
Best Local Similarity: 39.71% Mismatches: 33
Query Match: 12.36% Indels: 0
DB: 1 Gaps: 0

US-09-847-539A-6 (1-159) x AP002553 (1-297816)
QY 50 AlaThrThrAlaLeuGluAlaAlaSerSerAspAlaLeuGluAlaLeuAlaAspGlnThr 69
DB 83902 GCTGATACTTCAGCAGAGGATGCATCGAGTCAGCCGCGCAGCGGAGAAAGTGCAGCC 83961
QY 70 AspAlaLeuGlnSerGluGluAlaAlaValValValValValValValValValValVal 89
DB 83962 TCTGCAAGAAGTCAAGAGGAGGCTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 84021
QY 90 LeuGluAlaLeuAlaAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaGluValValGln 109
DB 84022 AGTGAGTCATTTACAAAGTGCACAGATGCCGAGTTGTTCACAAAAGAGCGGCGGAGTGC 84081
QY 110 SerAspAlaAlaAlaSerAspAla 117
DB 84082 GCCGGTAATGCAGCCAGGATGCA 84105

RESULT 22
AL663060
LOCUS
DEFINITION
Mus musculus chromosome 11 clone RP23-239H6, ** SEQUENCING IN
PROGRESS ***, in unordered pieces.
ACCESSION
AL663060
VERSION
AL663060.2 GI:18135362
HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE
house mouse.
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Burton, J.
Direct Submission
Submitted (08-JAN-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk
On Jan 11, 2002 this sequence version replaced gi:18071022.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BM239H6
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of bases
Consensus quality: 218850 bases at least Q40
Consensus quality: 219288 bases at least Q30
Consensus quality: 219705 bases at least Q20
Insert size: 220120; sum-of-contigs
Insert size: 237848; 1.5% error; agarose-fp

```

```

CDS      complement(8203..9492)

Alignment Scores:
Pred. No.:      0.093      Length:      13501
Score:          96.00      Matches:      27
Percent Similarity: 51.47%      Conservative: 8
Best Local Similarity: 39.71%      Mismatches: 33
Query Match:     12.36%      Indels:      0
DB:              1          Gaps:      0

US-09-847-539a-6 (1-159) x AE005258 (1-13501)
Qy 50 AlaThrThraIleGluAlaSerSerAspAlaLeuGluAlaLeuAlaSpGlnThr 69
   ||| |||::: ||| |||::: ||| |||::: ||| |||::: ||| |||::: |||
Db 1169 GCTGATACTTCAGCAGGATGCTATCGAGTCAGCCCGCGGAGGAGAAAGTCAGCC 1228
   ||| |||::: ||| |||::: ||| |||::: ||| |||::: ||| |||::: |||
Qy 70 AspAlaLeuGlnSerGluAlaAlaValValLysAlaAspAsnAlaAlaSerAspAla 89
   ||| |||::: ||| |||::: ||| |||::: ||| |||::: ||| |||::: |||
Db 1229 TCTGCAAGAAGTCAGAGGAGCGTCCTGCTCAGCTCTGAGCGCGCTCAAAAAGCC 1288
   ||| |||::: ||| |||::: ||| |||::: ||| |||::: ||| |||::: |||
Qy 90 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluAlaAlaGluValValGln 109
   ||| |||::: ||| |||::: ||| |||::: ||| |||::: ||| |||::: |||
Db 1289 AGTGAGTCATTACAAAGTGCAACAGATGCCGAGTGTTGTCAAAAGACCGCAGAAAGTCCA 1348
   ||| |||::: ||| |||::: ||| |||::: ||| |||::: ||| |||::: |||
Qy 110 SerAspAsnAlaAlaSerAspAla 117
   ::| ||| |||::: ||| |||::: ||| |||::: ||| |||::: ||| |||::: |||
Db 1349 GCCGTAATGCAGCCAGGATGCA 1372
   ::| ||| |||::: ||| |||::: ||| |||::: ||| |||::: ||| |||::: |||

RESULT 21
AP002553      AP002553      297816 bp      DNA      linear      BCT 07-MAR-2001
LOCUS      Escherichia coli O157:H7 DNA, complete genome, section 4/20.
DEFINITION      Escherichia coli O157:H7 DNA, complete genome, section 4/20.
ACCESSION      AP002553 BA000007
VERSION      AP002553.1 GI:13360211
KEYWORDS
SOURCE      Escherichia coli O157:H7 (strain:O157:H7, sub_strain:RIMD 0509952)
DNA.
ORGANISM      Escherichia coli O157:H7
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE
AUTHORS      Makino,K., Yokoyama,K., Kubota,Y., Yutsudo,C.H., Kimura,S.,
Kurokawa,K., Ishii,K., Hattori,M., Tatsuno,I., Abe,H., Iida,T.,
Yamamoto,K., Ohnishi,M., Hayashi,T., Yasunaga,T., Honda,T.,
Sasakawa,C. and Shinagawa,H.
TITLE      Complete nucleotide sequence of the prophage VT2-Sakai carrying the
verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7
derived from the Sakai outbreak
Genes Genet. Syst. 74 (5), 227-239 (1999)
JOURNAL
MEDLINE
REFERENCE
AUTHORS      Ohnishi,M., Murata,T., Nakayama,K., Kuhara,S., Hattori,M.,
Kurokawa,K., Yasunaga,T., Yokoyama,K., Makino,K., Shinagawa,H. and
Hayashi,T.
TITLE      Comparative analysis of the whole set of rRNA operons between an
enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an
Escherichia coli K-12 strain MG1655
Syst. Appl. Microbiol. 23 (3), 315-324 (2000)
JOURNAL
MEDLINE
REFERENCE
AUTHORS      Yokoyama,K., Makino,K., Kubota,Y., Watanabe,M., Kimura,S.,
Yutsudo,C.H., Kurokawa,K., Ishii,K., Hattori,M., Abe,H., Iida,T.,
Yamamoto,K., Hayashi,T., Yasunaga,T., Honda,T., Sasakawa,C. and
Shinagawa,H.
TITLE      Complete nucleotide sequence of the prophage VT1-Sakai carrying the
Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli
O157:H7 strain derived from the Sakai outbreak
Gene 258 (1-2), 127-139 (2000)
JOURNAL
MEDLINE
REFERENCE
AUTHORS      Hayashi,T., Makino,K., Ohnishi,M., Kurokawa,K., Ishii,K.,
Yokoyama,K., Han,C.-G., Ohtsubo,E., Nakayama,K., Murata,T.,
Tanaka,M., Tobe,T., Iida,T., Takami,H., Honda,T., Sasakawa,C.,
Ogasawara,N., Yasunaga,T., Kuhara,S., Shiba,T., Hattori,M. and
Shinagawa,H.
TITLE      Complete genome sequence of enterohemorrhagic Escherichia coli
O157:H7 and genomic comparison with a laboratory strain K-12
DNA Res. 8 (1), 11-22 (2001)
JOURNAL
MEDLINE
REFERENCE
AUTHORS      Ohnishi,M., Kurokawa,K., Makino,K., Yasunaga,T., Shinagawa,H. and
Hayashi,T.
TITLE      Direct Submission
JOURNAL
Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome
Information Research Center; 3-1, Yamadaoka, Suita, Osaka 565-0871,
Japan (E-mail:ken@gen-info.osaka-u.ac.jp,
URL:http://www.gen-info.osaka-u.ac.jp/, Tel:81-6-6879-8365,
Fax:81-6-6879-2047)
genome project.
Location/Qualifiers
1..297816
/organism="Escherichia coli O157:H7"
/strain="O157:H7"
/sub_strain="RIMD 0509952"
/db_xref="taxon:83334"
79..1245
/gene="ECS0753"
79..1245
/gene="ECS0753"
/feature="ECS0753"
/feature="similar to SUCC_ECOLI gi1786948 percent identity
100 in 388 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="succinyl-CoA synthetase beta subunit"
/protein_id="BAB34176.1"
/db_xref="GI:13360212"
/translation="MNLHEYCAKOLFARYGLPAPGVACTTPREAEAAKIGAGPWV
VKQVHAGKAGKAGGVKVVNSKEDIRAFENWLGKRLVYQTDTANGDPVQILVEAT
DIAKELYLGAVDNRVSRVFMASTEGVEIEKVAETPHLIHKVALDPLTGMPYOG
RELAFKLGLEKLVQOFTKIFMGLATIFLERDLALIEINPLVITKQDGLCLDGKLLG
DGNALFPDRLRMROSDQEDPRQAQAWELNYVALDNGICMVGAGLMTGMDIV
KLHGEPANFLDVGGGATKERVTEAFKIIISDDKVKAVLVNIFGGIVRCDLIADGIIG
AAVAGVNVFVPPVLEGNNAELGAKKLADSLNIIAAKGLTDAQQVVAVEGK"
1245..2114
/gene="ECS0754"
/feature="ECS0754"
/feature="similar to SUCD_ECOLI gi1786949 percent identity
100 in 289 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="succinyl-CoA synthetase alpha subunit"
/protein_id="BAB34177.1"
/db_xref="GI:13360213"
/translation="MSILIDKNTKVICQFTGSGTFSQAIAYGTMVGVPYPGKG
GTHLGLFVNTVREAVAGATASVIYVPAPPKDSILEIDAIGIKLIITIGIPT
LDMLTVKLDDEAGRMIGPCVITPGECKIGIQPHHKPKGVIGVSRGSLTTE
AVKQTTDYGQSTCVGIGDPIPGSNFIDILEFEDKDPQTEATVIMGEIGSGAEAA
AAYIKHVTKPVVYIAGVTAPKGRMHAGAIAGGKTADERFAALEAAGVKTVERS
LADIGEALKTVLK"
2781..3686
/gene="ECS0755"
/feature="ECS0755"
/feature="ECS0755"
/feature="probable transcriptional regulator,similar to
transcriptional regulators e.g. glycine cleavage system
transcription activator (gcv operon activator) -
Escherichia coli gi147043|sp|P32064|GCVA_ECOLI percent
identity 31 in 300 aa"
/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="putative transcription regulator"
/protein_id="BAB34178.1"

```

```

[prophage P-Eiba]}
/codon_start=1
/transl_table=11
/product="putative outer membrane protein of prophage
CP-933k"
/protein_id="AAG55137.1"
/db_xref="GI:12513752"
/translation="MRKVCAAITLSAICLAVSGVPAWASEHQSTLSAGYLHASTDAPG
SDLLNGINVKYIYETDILGLITSESYANAEDQKTHYSIDTRWHDYVVRNRFVSMAG
PSVRVNEFSATAMAGVAYSRVSTFSGDYFRVTDNKRKTHDLVLTGSDDARYSNTSLAW
GAGVQNFTEVSADVAVAVEXSGSDWRDGTDFIVGVYK"
719..2041
/gene="20982"
CDS
719..2041
/function="20982"
/feature="putative structure; Structural component (Phage
or Prophage Related)"
/notes="Residues 164 to 440 are 68.79 pct identical
to residues 381 to 645 from GenPept 118 :
g114585436|gb|AAD25464.1|AF125520_59 (AF125520) putative
tail fiber protein [Bacteriophage 933W]"
/codon_start=1
/transl_table=11
/product="putative tail component of prophage CP-933k"
/protein_id="AAG55138.1"
/db_xref="GI:12513753"
/translation="MGMAVQISGVLKDGAGPIQNCITQLKARNSTXVVNTVASE
NDEAGRYSMDVEYGOYSVTLVVEGPPSHAGTISYVEDSQPTLNDFIGANTDODAR
PEALRFQOMVEAEARHAEAEKKNAGEAETSARNAGISAKSAEASANAADTSAEDASE
SARQAESAASAEASAEASAEASAEASAEASAEASAEASAEASAEASAEASAEASAE
TSTKRAESASAEASAEASAEASAEASAEASAEASAEASAEASAEASAEASAEASAE
RGDTAGATGRTGRTGRTGRTGRTGRTGRTGRTGRTGRTGRTGRTGRTGRTGRTGRT
GRTGRTGRTGRTGRTGRTGRTGRTGRTGRTGRTGRTGRTGRTGRTGRTGRTGRTGRT
TGLTLDPADATQVOMFQHLQVRFEGDQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV
2004..2312
/gene="20984"
CDS
2004..2312
/gene="20984"
/function="20984"
/feature="orf; Other or unknown (Phage or Prophage
Related)"
/notes="Residues 1 to 102 of 102 are 98.03 pct identical to
residues 28 to 129 from GenPept 118 :
g114585437|gb|AAD25465.1|AF125520_60 (AF125520)
hypothetical protein [Bacteriophage 933W]"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933k"
/protein_id="AAG55139.1"
/db_xref="GI:12513754"
/translation="MKWAVIOAELENDMNLRLKQSLCCGCKHDDCNGRSITQALR
LGPADILESDENGIIPEQDRVITQVYILDADKQIQCVVVRPLQILRADGTWENIGMK
"
2489..3469
/gene="20985"
CDS
2489..3469
/gene="20985"
/function="20985"
/feature="orf; Other or unknown (Phage or Prophage
Related)"
/notes="Residues 13 to 326 of 326 are 55.73 pct identical
to residues 18 to 331 of 336 from GenPept 118 :
g116960367|gb|AAF33527.1| (AF170176) hypothetical protein
predicted by Glimmer [Salmonella typhimurium LT2]"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933k"
/protein_id="AAG55140.1"
/db_xref="GI:12513755"
/translation="MSPRTTFHNSVNIQSSPQCVTSFACKVELKVIDEKTPIILF
QNFENPERKKEDEPIVNTKQHPYLDNVTNAIRESDMIGIFVGDPSVOKRTAFS
KLERFENMIIYREDVDFSMYDRKLSIYHDIIICEORLTKEDKRDYLLNLEKELR
EISKAQDSLISYAKRNHAWDFRNALLAKAGETFRCTYNTKNHGISFGGCTYLD
MDMLITGLTITAPDGIISMHVDRNDSVNIENSALIVNRSNHPALLEGISPMHMKVD
AHPYVDGLGKGVKYFNFPTPLNHNHFCDFIEFNHPIIMNTSQYTCSSW"

```

```

gene
CDS
3530..4522
/gene="20986"
3530..4522
/gene="20986"
/function="20986"
/feature="orf; Other or unknown (Phage or Prophage
Related)"
/notes="No significant matches"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933k"
/protein_id="AAG55141.1"
/db_xref="GI:12513756"
/translation="MKIPSLQSNFSAFAGYSAPIAPNRAENAYADYVLDIGKRIPL
SAADLSNVYESVIRAVHDSRSLIQHTVDMIGNTVLDALSRQSFDRDAYSVGIHNEK
VHIGCIKYNEYELNEESSVKIDDIQSLTCNELYEDYGOEPIFFICEAGENDNEEPPY
VFSVAPDSDSYEMPMSOQELIHEIITHVTSSDPSGDSNIELGPTTEILARRVAQELG
WSPVPFKGYAEPEAREAHRLRLNALRQAAREHEERAFERLGTISDRYEASPDFT
EYSVSNIGYGFIIQHDFFGLAINDLQDANQIQLYHGAPYITFTFGDVKHNOQ"
5350..6231
/gene="20989"
CDS
5350..6231
/gene="20989"
/function="20989"
/feature="orf; Other or unknown (Phage or Prophage
Related)"
/notes="No significant matches"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933k"
/protein_id="AAG55142.1"
/db_xref="GI:12513757"
/translation="MLSPYSVNLGCSWNSLTRNLTPDNRLVSSVRDAVHSDNGAQV
KVGNTRYRVATDNPVDPVAPPPELPSVDYVNLSPGVNPKGKCEISKEIKYELMLNADSPYSAM
ERGIYHSNRPLDPPVAPPPELPSVDYVNLSPGVNPKGKCEISKEIKYELMLNADSPYSAM
FTTSQNEEVTSEVRCTNOYAGSAEKIYGNNGDIIIGIRMDKINGESLNLISLSPAQ
AEHAIYDMFDRLEQKGLFVDTTETNVLVDRAKNENFPIDISSYNSVSDRSWSESQIMQ
SYHGGKQDLISVVLSKI"
6462..7160
/gene="20990"
CDS
6462..7160
/gene="20990"
/function="20990"
/feature="orf; Other or unknown (Phage or Prophage
Related)"
/notes="No significant matches"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933k"
/protein_id="AAG55143.1"
/db_xref="GI:12513758"
/translation="MRPTSLNLVLHQSRSWSDTDIESLVKASSVOMIKNNPOLRF
QGTDRNIYQOIEAALDKIGTETGRVLLNAIESISRLKSETVVIHLNLSRLGVMHARD
IDAENHRGTGSDPHCNLNAVVEYPCGEGISVDPFHATIVFHELLHVFHNLERLKVES
SRAESOKYSPLLLEARTVGLGAFSEEVLENKFHEIGMPRTSYPPXDSALJHDDNT
VSLGFOQVRLHPLL"
complement(7668..8144)
/gene="20992"
/feature="20992"
/feature="orf; Other or unknown (Phage or Prophage
Related)"
/notes="Residues 1 to 158 of 158 are 99.36 pct identical to
residues 1 to 158 of 158 from Escherichia coli K-12 Strain
MG1655: B0773"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAG55144.1"
/db_xref="GI:12513759"
/translation="MKLISNDLRDGDKLPHRVHFGMGVYDGNISPHLADDDVPACTK
SVFVTCYDPAFTGSGMHWVYVNLNPAIDRVLPQCGSLVAMPDGLVLTQTRDFQKTK
YDGAAPPKGETHYRFTVHALDVERIDVDGASGAMGVFNHFIHSLASASITAMFS"
complement(8203..9492)
/gene="20993"
/feature="20993"

```







```

/product="putative periplasmic protein"
/protein_id="CAD09869.1"
/db_xref="GI:16505983"
/translation="MKTVLDTLLLVIAFVLCIFGAIKALAPVNGAMVYDSGEO
FYVYDRVLSSPYLTKRNGKLSGVGDKFYVNGSGDVGVGHAKNASYLFDCKEYG"
complement(36..43)
/gene="HCM2.0002c"
/note="possible RBS"
/feature
  repeat_unit
    repeat 1: 57 bp repeat
  gene
    complement(1647..2489)
  CDS
    complement(1647..2489)
    /gene="HCM2.0006c"
    /note="HCM2.0006c, possible lipoprotein, len: 280 aa;
    shows weak similarity to TR:AAF10061 (EMBL:AE001907),
    DR0482, Deinococcus radiodurans B-cell receptor associated
    protein-related protein (328 aa), fasta scores; E():
    4.9e-07, 23.4% identity in 265 aa overlap. Highly similar
    to TR:CAB55231 (EMBL:AL117211), YPMT1.49C, Versinia pestis
    CO-92 putative lipoprotein from plasmid pMT1 (276 aa),
    fasta scores; E(): 0, 97.5% identity in 280 aa overlap and
    TR:O68763 (EMBL:AF074611), Y1114, Versinia pestis KIM5
    hypothetical protein from plasmid pMT1 (276 aa), fasta
    scores; E(): 0, 97.5% identity in 280 aa overlap. Contains
    a possible N-terminal signal sequence and an appropriately
    positioned PS00013 Prokaryotic membrane lipoprotein lipid
    attachment site"
    /codon_start=1
    /transl_table=11
    /product="putative lipoprotein"
    /protein_id="CAD09873.1"
    /db_xref="GI:16505987"
    /translation="MKKGLLAVALTAICTMGLTGCDRVEPYGVKIKLGEDKIGIE
    VVGVRGWTGLNTEYTFQMKTYDEPFTFQMSDGTGAIHGKIGVYLVNRDKVTT
    VFQYRGKGVDDITDILRKIADSLNRLASMTTDFIDGKASLLNALDKDIAQEMS
    PVGIEVLSLVGKPKDPDYVIESINAKVTANOKTLQRCQVEQKAEANMLRQAEQ
    EAGEADAIRKRAQAEADAIKLCEALRQNPVNELEAINKNWQQLPYQWTEGANTFF
    IALK"
    complement(2427..2459)
    /gene="HCM2.0006c"
    /note="PS00013 Prokaryotic membrane lipoprotein lipid
    attachment site"
    complement(2493..2693)
    /gene="HCM2.0007c"
    /note="HCM2.0007c"
    highly similar to TR:CAB62372 (EMBL:AL117211), YPMT1.49AC,
    Versinia pestis CO-92 hypothetical protein from plasmid
    pMT1 (66 aa), fasta scores; E(): 3.5e-25, 97.0% identity
    in 66 aa overlap. Contains hydrophobic, possible
    Alignment Scores:
    Pred. No.: 0.0807 Length: 106516
    Score: 104.00 Matches: 28
    Percent Similarity: 49.38% Conservative: 12
    Best Local Similarity: 34.57% Mismatches: 41
    Query Match: 13.38% Indels: 0
    DB: 1 Gaps: 0
    US-09-847-539a-6 (1-159) x STYPPHCH2 (1-106516)
    Qy 49 GluAlaThrAlaIleGluAlaAlaSerSerAspAlaLeuGluAlaLeuAlaAspGln 68
    ||||| ::::::::::::::: ||| ||| ::
    Db 39715 GAGCGTTAACTCTAAACAAGCAGCTCGCGGACGGAAGCTAATGCTAAGGCAAGTCAA 39656
    ||||| ::::::::::::::: ||| ||| ::
    Qy 69 ThrAspAlaLeuGlnSerGluAlaAlaValValLysAlaAspAlaAlaSerAsp 88
    ||| ::::::::::::::: ||||| ::
    Db 39655 AACGCTGAGCTGCCTCTCAGCAGCAGCGGCTACCAAGTAAAGTAAAGCCGCGGACAGC 39596
    ||| ||| ||||| ||||| ||||| ||||| ::
    Qy 89 AlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaValVal 108
    ||| ||| ||||| ||||| ||||| ||||| ::
    Db 39595 AAAGAGCTGCCGCGACATCAACAAACAGCTGCATTACAAAGTGACGAGTGGCAGCCGCC 39536

```

gene	/rpt_unit=612. .622 612. .770 /gene="EBNA1"				VERSION KEYWORDS SOURCE ORGANISM
CDS	1843. .2259 /note="potential 15kda protein; similar to BKRF2 of EBV, Swiss-Prot Accession Number P03212"				AL513384.1 GI:16505981 Salmonella enterica subsp. enterica serovar Typhi. Salmonella enterica subsp. enterica serovar Typhi Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Salmonella.
CDS	/codon_start=1 /protein_id="AAA66374.1" /db_xref="GI:775216" /translation="MHAARFVSLVALFVLPWTGWAYPCCHVTQLSQOHHLALE NISDIYLSNKNKCDGFSLASLSPRNSQLVISRCANGLNVASTFIAVLRKRSALT SQLELLNTLSLEYASFVSFVEDLFGANLNRYAHHAG" 2241. .3008 /note="similar to BKRF3 of EBV encoded by GenBank Accession Number V01555; similar to EBV uracil DNA glycosylase, Swiss-Prot Accession Number P12888"				1 (bases 1 to 106516) Parkhill,J., Dougan,G., James,K.D., Thomson,N.R., Pickard,D., Wain,J., Churcher,C., Mungall,K.L., Bentley,S.D., Holden,M.T.G., Sebaihia,M., Baker,S., Basham,D., Brooks,K., Chillingworth,T.G., Connerton,P., Cronin,A., Davis,P., Davies,K.M., Dowd,L., White,N., Farrar,J., Feltwell,T., Hamlin,N., Haque,A., Hien,T.T., Holroyd,S., Jagels,K., Krogh,A., Larsen,T.S., Leather,S., Moule,S., O'Gaora,P., Parry,C., Quail,M., Rutherford,K., Simmonds,M., Skelton,J., Stevens,K., Whitehead,S. and Barrell,B.G. Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhi CT18 Nature 413 (6858), 848-852 (2001)
CDS	/citation={1} /codon_start=1 /protein_id="AAA66375.1" /db_xref="GI:775217" /translation="NASRGLDLWLDEHVWKKEGVKGENLLLPDPWLNFLQLSPIF QRKLAALACIRRLTQATIIYPEEDMCNARFCDPSDIKVVILGDDPHYGGGAGLA FSVADGFVPPSLRIYAEIHSRSLPEFSPDHGCLDANASQGVLLNTLTIVQKGP SHADIGWAFTHDIIIALISERLKACVFMLWGAAGDKASLINSKRHLVLTQHPSPLA QNSTKSAQQKEVGNHFEVLANNFLREKGLGEIDWRL" 3019. .>3222 /note="similar to BKRF4 of EBV, Swiss-Prot Accession Number P30117"				2 (bases 1 to 106516) Parkhill,J. Direct Submission Submitted (25-OCT-2001) Submitted on behalf of the Salmonella sequencing team, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk
BASE COUNT	727 a 857 c 1035 g 603 t				Notes: Details of S. typhi sequencing at the Sanger Centre are available on the World Wide Web. (URL, http://www.sanger.ac.uk/Projects/S_typhi/).
ORIGIN	EVVSTESDYSEEDIDEEDF				Location/Qualifiers 1. .106516 /organism="Salmonella enterica subsp. enterica serovar Typhi"
Alignment Scores:					/plasmid="pHCM2" /strain="CT18"
Pred. No.:	0.00187	Length:	3222		/db_xref="taxon:90370"
Score:	104.00	Matches:	24		complement(1. .29) /gene="HCM2.0001c"
Percent Similarity:	56.47%	Conservative:	24		complement(1. .29) /partial
Best Local Similarity:	28.24%	Mismatches:	37		/gene="HCM2.0001c"
Query Match:	13.38%	Indels:	0		/note="HCM2.0001c", hypothetical protein, len: 68 aa; highly similar to TR:CA855227 (EMBL:AL117211), YPWT1.45C, Yersinia pestis CO-92 hypothetical protein from plasmid pMT1 (71 aa), fasta scores: E(): 5.9e-25, 93.3% identity in 60 aa overlap and TR:O88760 (EMBL:AF074611), Y1119, Yersinia pestis KIM5 hypothetical protein from plasmid pMT1 (71 aa), fasta scores: E(): 5.9e-25, 93.3% identity in 60 aa overlap. Spans the end of the sequence"
DB:	14	Gaps:	0		/codon_start=1 /transl_table=11 /product="hypothetical protein"
US-09-847-539A-6 (1-159) x HPU23857 (1-3222)					/protein_id="CAD09868.1" /db_xref="GI:16505982" /translation="MAVTLAGLE"
QY 63 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLysAla 82	::: ::    ::   ::    ::          :::				1. .232 /note="95.7% identical to Yersinia pestis CO-92 plasmid pMT1 (EMBL:AL117211) bases 48304. .48535"
Db 616 GAAGCGGTGCAGGAGGAGGAGCGGTGCAGGAGGAGCGGTGCAGGAGGAG 675	:::     ::::      ::: ::    ::   ::				complement(29. .334) /gene="HCM2.0002c"
QY 83 AspAsnAlaAlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSer 102	:::     ::::      ::: ::    ::   ::				complement(29. .334) /gene="HCM2.0002c"
Db 676 GAGGAAGCGGTGCAGGAGGAGGAGGAGCGGTGCAGGAGGAGGAGCGGTGCAGGAG 735	:::     ::::      ::: ::    ::   ::				/note="HCM2.0002c", possible periplasmic protein, len: 101 aa; similar to TR:CA855228 (EMBL:AL117211), YPWT1.46C, Yersinia pestis hypothetical protein from plasmid pMT1 (111 aa), fasta scores: E(): 2.8e-26, 95.5% identity in 67 aa overlap. The N-terminal 37 aa are highly similar to YPWT1.46AC (EMBL:AL117211) Yersinia pestis hypothetical protein from plasmid pMT1 (59 aa) (86.8% identity in 38 aa overlap) which lies immediately upstream of YPWT1.46C in pMT1. Contains a possible N-terminal signal sequence"
QY 103 GluGluAlaGluValValGlnSerAspAsnAlaAlaSerAspAlaTrpGluLysAlaAla 122	:::         ::::      :::     ::::      :::				/codon_start=1 /transl_table=11
Db 736 CAGAGGAACCGGTGCAGGAGGAGGAGGAGCGGTGCAGTCGAGGAGGAGGAGCA 795	:::     ::::      ::: ::    ::   ::				
QY 123 ThrProileAlaLeuAspValLysLysThrLysAspThrLysProValValLysLysGlu 142	:::    :::    :::    :::    :::				
Db 796 GAGGGACAGCAGTAGAGGTAGAGGAAGGAGGAGGAGGAGGAGGAGGAGGAG 855	:::    :::    :::    :::    :::				
QY 143 GluArgGluAsnVal 147	:::    :::    :::    :::    :::				
Db 856 GTCAGGAGGAACATA 870	:::    :::    :::    :::    :::				
RESULT 18					
STYPPHCM2/c					
LOCUS					
DEFINITION					
Salmonella enterica serovar Typhi (Salmonella typhi) strain CT18					
plasmid pHCM2.					
AL513384					
ACCESSION					



----- Summary Statistics  
 Center project name: H\_NH0461M18  
 -----

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

#### MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

#### SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tateno, M., Catanese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBACe3.6

#### NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-341M1, 2000 bp overlap; the clone sequenced to the right is RP11-289K3. Actual start of this clone is at base position 162181 of RP11-341M1; actual end is at base position 167227 of RP11-461M18.

Data from AC016725 and AC017031 was used to finish this clone, AC020602. There is a tandem repeat from base position 33099 to 33933.

#### FEATURES

source

Location/Qualifiers  
 1. .167227  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="2"  
 /map="2"  
 /clone="RP11-461M18"  
 /clone\_lib="RPCI-11"  
 /rpt\_family="MIR"  
 250. .316  
 /rpt\_family="MIR"  
 878. .923  
 /rpt\_family="MIR"  
 1376. .1470  
 /rpt\_family="MIR"  
 2361. .2391  
 /rpt\_family="(TTTA)n"  
 2850. .3229  
 /rpt\_family="MALR"  
 3260. .3347  
 /rpt\_family="ERVL"  
 3488. .3516  
 /rpt\_family="(TTG)n"  
 3490. .3796  
 /rpt\_family="Alu"  
 5077. .5106  
 /rpt\_family="AT\_rich"  
 5090. .5289  
 /rpt\_family="L1"  
 6473. .6782  
 /rpt\_family="Alu"  
 6754. .6830

repeat\_region  
 6841. .7120  
 /rpt\_family="GA-rich"  
 /rpt\_family="Alu"  
 7235. .7580  
 /rpt\_family="L1"  
 7792. .7948  
 /rpt\_family="L1"  
 8071. .8364  
 /rpt\_family="Alu"  
 8486. .8631  
 /rpt\_family="L1"  
 8668. .8693  
 /rpt\_family="AT\_rich"  
 8738. .8887  
 /rpt\_family="Alu"  
 8891. .9208  
 /rpt\_family="Alu"  
 9061. .9090  
 /rpt\_family="AT\_rich"  
 9210. .9343  
 /rpt\_family="L1"  
 9649. .9701  
 /rpt\_family="MIR"  
 9702. .9830  
 /rpt\_family="T2\_type"  
 9831. .9992  
 /rpt\_family="MIR"  
 10568. .10588  
 /rpt\_family="AT\_rich"  
 11202. .11499  
 /rpt\_family="Alu"  
 11531. .11609  
 /rpt\_family="Alu"  
 11612. .11641  
 /rpt\_family="(TTG)n"  
 11619. .11792  
 /rpt\_family="Alu"  
 11793. .11850  
 /rpt\_family="Alu"  
 12082. .12261  
 /rpt\_family="MER1\_type"  
 12485. .12870  
 /rpt\_family="L1"  
 12890. .12924  
 /rpt\_family="(TTG)n"  
 12921. .12982  
 /rpt\_family="L1"  
 12983. .13285  
 /rpt\_family="Alu"  
 13257. .13292  
 /rpt\_family="(GAA)n"  
 13286. .13552  
 /rpt\_family="L1"  
 13726. .14016  
 /rpt\_family="L1"  
 14020. .14331  
 /rpt\_family="Alu"  
 14335. .14439  
 /rpt\_family="Alu"  
 14509. .15557  
 /rpt\_family="L1"  
 15558. .15850  
 /rpt\_family="Alu"  
 15851. .15904  
 /rpt\_family="L1"  
 15914. .16150  
 /rpt\_family="Alu"  
 16156. .16302  
 /rpt\_family="L1"  
 16313. .16664  
 /rpt\_family="L1"  
 16682. .16939  
 /rpt\_family="L1"

## Alignment Scores:

Pred. No.: 1.13e-08 Length: 1576  
 Score: 142.00 Matches: 37  
 Percent Similarity: 60.61% Conservatives: 3  
 Best Local Similarity: 56.06% Mismatches: 26  
 Query Match: 18.28% Indels: 1  
 DB: 1 Gaps: 0

US-09-847-539A-6 (1-159) x SG148PG (1-1576)

Qy 1 ValAspSerProIleGluGlnProArgIleIleProAsnGlyGlyThrLeuThrAsnLeu 20  
 |||||  
 Db 391 GTTGACTCACCAGATACCCCAATTATTCGTAATGGTGTGAATTAACATAATCTT 450  
 |||||  
 Qy 21 LeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaIleAspGluLeu 40  
 |||||  
 Db 451 CTGGGAATTCAGACACACATCTTGGCTTGAATGAAGAGAGTGCTACAGCTGATTGTG 510  
 |||||  
 Qy 41 LysLysGlnAlaIleGluAspLysGluAlaThrThrAlaIleGluAlaAlaSerSerAsp 60  
 |||||  
 Db 511 ACAGCAGCAGCGGTAGCGGATCTGTGCAGCAGCGGAGTGAATGCTGGGCAGCA- 569  
 |||||  
 Qy 61 AlaLeuGluAlaLeuAla 66  
 |||||  
 Db 570 GCTTGGGAAGCAGCGGCA 587

## RESULT 14

STRABP 969 bp DNA linear BCT 26-APR-1993  
 LOCUS Streptococcus canis (group G) albumin-binding protein gene, partial cds.  
 DEFINITION M95520.1 GI:153554  
 ACCESSION albumin-binding protein.  
 VERSION Streptococcus canis (group G) (individual isolate DG12) DNA.  
 KEYWORDS Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 SOURCE Streptococcus.  
 ORGANISM Streptococcus.  
 REFERENCE 1 (bases 1 to 969)  
 AUTHORS Sjobring,U.  
 TITLE Isolation and molecular characterization of a novel albumin-binding protein from group G streptococci  
 JOURNAL Infect. Immun. 60, 3601-3608 (1992)  
 MEDLINE 92363555  
 FEATURES Location/Qualifiers  
 source 1..969  
 /organism="Streptococcus canis"  
 /isolate="DG12"  
 /db\_xref="taxon:1329"  
 <1..>969  
 /codon\_start=1  
 /transl\_table=11  
 /product="albumin-binding protein"  
 /protein\_id="AAA6847.1"  
 /db\_xref="GI:153554"

## CDS

359 a 168 c 219 g 223 t  
 BASE COUNT  
 ORIGIN  
 /translation="SVLVGANTVSAQVTTTAAQAAKARETALASIAETESSIKKAQLD  
 SVKEIKASVDREITELSAQADQIVSAQADNEAITRAEDSSKAAVAAQAAQANTAKA  
 EADELAKESSDAWEKAALDOAKAALKEFDYGVSNYKNIINKAKTVEGIMEL  
 QAQVESAKKARISEATDGLSKQTSADTLKSLSEAKEMAIRELDAQVSDF  
 YKNNKNAKTVEGVVLDLNLNSLAEPEAKPEVKPEVKPEKPAEVRKPEVKPEV  
 KPEVKPEVKPEVKPEVKPEVKPEKPAEVRKPEVKPEVKPEVKPEVKPEVKPEV"

## Alignment Scores:

Pred. No.: 7.62e-07 Length: 969  
 Score: 126.00 Matches: 26  
 Percent Similarity: 65.57% Conservatives: 14  
 Best Local Similarity: 42.62% Mismatches: 21  
 Query Match: 16.22% Indels: 0  
 DB: 1 Gaps: 0

US-09-847-539A-6 (1-159) x STRABP (1-969)

Qy 62 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluAlaAlaValVallys 81  
 |||||  
 Db 184 TTGTGCGCGCAGCAGACCATGTTTCTTCACCAAGCAGATAACGAGCTATTACTAAG 243  
 |||||  
 Qy 82 AlaAspAsnAlaAlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGln 101  
 |||||  
 Db 244 GCAGAAAGAGATTCTTCCCAAGCATGGGAAGCTGCTGCGGATCAAGCAATACTGCTAAA 303  
 |||||  
 Qy 102 SerGluGluAlaGluValValGlnSerAspAsnAlaAlaSerAspAlaTrpGluLysAla 121  
 |||||  
 Db 304 GCAGAACAGATGAACCTTGCACAAAGCAGAGAAAAGAATCATCAGATGCTTGGGAGAGGCT 363  
 |||||  
 Qy 122 Ala 122  
 |||||  
 Db 364 GCT 366

## RESULT 15

AC020602/c 167227 bp DNA linear PRI 10-JAN-2002  
 LOCUS Homo sapiens BAC clone RP11-461M18 from 2, complete sequence.  
 DEFINITION AC020602  
 ACCESSION AC020602.6 GI:17386508  
 VERSION HTG.  
 KEYWORDS human.  
 SOURCE Homo sapiens  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE 1 (bases 1 to 167227)  
 AUTHORS Sulston,J.E. and Waterston,R.  
 TITLE Toward a complete human genome sequence  
 JOURNAL Genome Res. 8 (11), 1097-1108 (1998)  
 MEDLINE 99063792  
 REFERENCE 2 (bases 1 to 167227)  
 AUTHORS Waligorski,J., Cotton,M. and Doeber,A.  
 TITLE The sequence of Homo sapiens BAC clone RP11-461M18  
 JOURNAL Unpublished (2002)  
 REFERENCE 3 (bases 1 to 167227)  
 AUTHORS Waterston,R.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (05-JAN-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

REFERENCE 4 (bases 1 to 167227)  
 AUTHORS Waterston,R.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (06-DEC-2001) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

REFERENCE 5 (bases 1 to 167227)  
 AUTHORS Waterston,R.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (07-DEC-2001) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

REFERENCE 6 (bases 1 to 167227)  
 AUTHORS Waterston,R.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (03-JAN-2002) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

REFERENCE 7 (bases 1 to 167227)  
 AUTHORS Waterston,R.  
 TITLE Direct Submission  
 JOURNAL Submitted (10-JAN-2002) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
 On Dec 6, 2001 this sequence version replaced gi:14029092.

COMMENT  
 ----- Genome Center  
 Center: Washington University Genome Sequencing Center  
 Center code: WUGSC  
 Web site: http://genome.wustl.edu/gsc  
 Contact: sapiens@watson.wustl.edu

```

repeat_region 1252..1416
/note="repetitious region 3"
repeat_region 1417..1461
/note="repetitious region 4"
repeat_region 1462..1626
/note="repetitious region 3"
repeat_region 1627..1671
/note="repetitious region 4"
repeat_region 1672..1836
/note="repetitious region 3"
misc_feature 1837..2046
/note="cell-wall spanning region"
misc_feature 2047..2124
/note="membrane spanning region"
repeat_unit 2157..2165
/note="inverted repeat A"
repeat_unit 2172..2180
/note="inverted repeat A"
BASE COUNT 907 a 412 c 491 g 646 t
ORIGIN

```

```

Alignment Scores:
Pred. No.: 3,9e-10 Length: 2456
Score: 155.00 Matches: 37
Percent Similarity: 62.12% Conservative: 4
Best Local Similarity: 56.06% Mismatches: 25
Query Match: 19.95% Indels: 0
DB: 1 Gaps: 0

```

US-09-847-539A-6 (1-159) x SGPRTG (1-2456)

```

QY 1 ValAspSerProTleGluGlnProArgIleIleProAsnGlyGlyThrLeuThrAsnLeu 20
|||||
Db 445 GTTGACTCACCACATCGAAGATACCCCAATATTTCGTAATGTTGGTGAATTAATCTT 504
|||||

QY 21 LeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaIleAspGluLeu 40
|||||
Db 505 CTGGGGAATTCAGAGACACACTGGCTTTCGTAATGAAGAGAGTGCTACAGCTGATTG 564
|||||

QY 41 LysLysGlnAlaIleGluAspLysGluAlaThrAlaIleGluAlaLeuSerSerAsp 60
|||||
Db 565 ACAGCAGCAGCGGTGCGGATCTGTGGCAGCAGCGGCGCTGAAAATGCTGGGCAGCA 624
|||||

QY 61 AlaLeuGluAlaLeuAla 66
|||
Db 625 GCTTGGGAAGCAGCGGCA 642
|||

```

```

RESULT 12
I08536
LOCUS I08536 1950 bp linear PAT 02-DEC-1994
DEFINITION Sequence 2 from Patent WO 8705025.
ACCESSION I08536
VERSION I08536.1 GI:588754
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1950)
AUTHORS Fahnstock,S.R.
TITLE CLONED STREPTOCOCCAL GENES ENCODING PROTEIN G AND THEIR USE TO
CONSTRUCT RECOMBINANT MICROORGANISMS TO PRODUCE PROTEIN G
JOURNAL Patent: WO 8705025-A 2 27-AUG-1987;
FEATURES
source
location/Qualifiers
1..1950
/organism="unknown"
BASE COUNT 706 a 324 c 397 g 523 t
ORIGIN

```

```

Alignment Scores:
Pred. No.: 5.5e-10 Length: 1950
Score: 153.00 Matches: 37
Percent Similarity: 62.12% Conservative: 4
Best Local Similarity: 56.06% Mismatches: 25

```

```

Query Match: 19.69% Indels: 0
DB: 6 Caps: 0
US-09-847-539A-6 (1-159) x I08536 (1-1950)
QY 1 ValAspSerProTleGluGlnProArgIleIleProAsnGlyGlyThrLeuThrAsnLeu 20
|||||
Db 677 GTTGACTCACCACATCGAAGATACCCCAATATTTCGTAATGTTGGTGAATTAATCTT 736
|||||

QY 21 LeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaIleAspGluLeu 40
|||||
Db 737 CTGGGGAATTCAGAGACACACTGGCTTTCGTAATGAAGAGAGTGCTACAGCTGATTG 796
|||||

QY 41 LysLysGlnAlaIleGluAspLysGluAlaThrAlaIleGluAlaLeuSerSerAsp 60
|||||
Db 797 ACAGCAGCAGCGGTGCGGATCTGTGGCAGCAGCGGCGCTGAAAATGCTGGGCAGCA 856
|||||

QY 61 AlaLeuGluAlaLeuAla 66
|||
Db 857 GCTTGGGAAGCAGCGGCA 874
|||

RESULT 13
SGI48PG
LOCUS SGI48PG 1576 bp DNA linear BCT 23-MAR-1993
DEFINITION Streptococcus G148 protein G' structural gene.
ACCESSION X53324
VERSION X53324.1 GI:288358
KEYWORDS Streptococcal protein G.
SOURCE Streptococcus sp. 'group G'.
ORGANISM Streptococcus sp. 'group G'.
Bacteria: Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
Streptococcus.
1 (bases 1 to 1576)
REFERENCE 1
AUTHORS Goward,C.R., Murphy,J.P., Atkinson,T. and Barstow,D.A.
TITLE Expression and purification of a truncated recombinant
Streptococcal protein G
JOURNAL Biochem. J. 267 (1), 171-177 (1990)
MEDLINE 90226312
FEATURES
source
location/Qualifiers
1..1576
/organism="Streptococcus sp. 'group G'"
/strain="G148"
/db_xref="taxon:1320"
/clone="SPG3"
292..1262
/gene="Protein G'gene"
292..603
/gene="Protein G'gene"
/note="Truncated gene"
/codon_start=1
/transl_table=11
/product="Protein G'"
/protein_id="CAA37409.1"
/db_xref="GI:288359"
/db_xref="SPTREMBL:Q54180"
/translation="MEKEKKVYFLKSAFGLASVSAAFVGVSTVPAVDSPIDETPII
RNGELTNLLGNSETTLAURNEESATADLTAAAVDTVAANAENAGQQLGKQRQQM
L"
705..1262
/gene="Protein G'gene"
/note="Truncated gene; Start codon TTG not ATG"
/codon_start=1
/transl_table=11
/product="Protein G'"
/protein_id="CAA37410.1"
/db_xref="GI:581662"
/db_xref="SPTREMBL:Q54181"
/translation="MKGETTTEAVDAATAEKVFKQYANDNGVDGTYDDATKTFVT
EKPEVIDASELTPAVTYYKLIVNGTKLGETTEAVDAATAEKVFKQYANDNGVDGEW
TYDDATKTFVTKEPEVIDASELTPAVTYYKLIVNGTKLGETTTKAVDAETAERKAFK
QYANDNGVDGVTYDDATKTFVTE"
BASE COUNT 572 a 273 c 327 g 404 t
ORIGIN

```

repeat\_region 1111..1221  
 /note="A2 repeated sequence"  
 repeat\_region 1354..1446  
 /note="A3 repeated sequence"  
 repeat\_region 1483..1647  
 /note="B1 repeated sequence"  
 repeat\_region 1499..1918  
 /note="210 bp duplication"  
 /rpt\_unit=1708..1918  
 repeat\_region 1693..1857  
 /note="B2 repeated sequence"  
 repeat\_region 1903..2067  
 /note="B3 repeated sequence"  
 BASE COUNT 868 a 408 c 482 g 626 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 3,78e-10 Length: 2384  
 Score: 155.00 Matches: 37  
 Percent Similarity: 62.12% Conservative: 4  
 Best Local Similarity: 56.06% Mismatches: 25  
 Query Match: 19.95% Indels: 0  
 DB: 1 Gaps: 0

US-09-847-539A-6 (1-159) x SGSPG (1-2384)

QY 1 ValAspSerProIleGluGlnProArgIleIleProAsnGlyGlyThrLeuThrAsnLeu 20  
 |||||  
 Db 676 GTTGACTCACCATCGAAGATACCCCAATATTTCGTAATGGTGGTGAATTAATCTT 735  
 QY 21 LeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaLeuAspGluLeu 40  
 |||||  
 Db 736 CTGGGAATTCAGACACAACACTGGCTTTCGTAATGAAGAGAGTGTACAGCTGATTG 795  
 QY 41 LysLysGlnAlaIleGluAspLysGluAlaThrAlaIleGluAlaAlaSerSerAsp 60  
 |||||  
 Db 796 ACAGCAGCAGCGGTAGCCGATCTGTGGCAGCAGCGGAGTGAATAATGCTGGGCGAGCA 855  
 QY 61 AlaLeuGluAlaLeuAla 66  
 |||||  
 Db 856 GCTTGGGAAGCAGCGGCA 873

RESULT 10  
 LOCUS 108537  
 DEFINITION Sequence 4 from Patent WO 8705025.  
 ACCESSION 108537  
 VERSION 108537.1 GI:588755  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 REFERENCE 1 (bases 1 to 2384)  
 AUTHORS Fahnestock,S.R.  
 TITLE CLONED STREPTOCOCCAL GENES ENCODING PROTEIN G AND THEIR USE TO  
 CONSTRUCT RECOMBINANT MICROORGANISMS TO PRODUCE PROTEIN G  
 JOURNAL Patent: WO 8705025-A 4 27-AUG-1987;  
 FEATURES Location/Qualifiers  
 source  
 1..2384  
 /organism="unknown"  
 BASE COUNT 868 a 410 c 480 g 626 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 3,78e-10 Length: 2384  
 Score: 155.00 Matches: 37  
 Percent Similarity: 62.12% Conservative: 4  
 Best Local Similarity: 56.06% Mismatches: 25  
 Query Match: 19.95% Indels: 0  
 DB: 6 Gaps: 0

US-09-847-539A-6 (1-159) x 108537 (1-2384)

QY 1 ValAspSerProIleGluGlnProArgIleIleProAsnGlyGlyThrLeuThrAsnLeu 20  
 |||||  
 Db 676 GTTGACTCACCATCGAAGATACCCCAATATTTCGTAATGGTGGTGAATTAATCTT 735  
 QY 21 LeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaLeuAspGluLeu 40  
 |||||  
 Db 736 CTGGGAATTCAGACACAACACTGGCTTTCGTAATGAAGAGAGTGTACAGCTGATTG 795  
 QY 41 LysLysGlnAlaIleGluAspLysGluAlaThrAlaIleGluAlaAlaSerSerAsp 60  
 |||||  
 Db 796 ACAGCAGCAGCGGTAGCCGATCTGTGGCAGCAGCGGAGTGAATAATGCTGGGCGAGCA 855  
 QY 61 AlaLeuGluAlaLeuAla 66  
 |||||  
 Db 856 GCTTGGGAAGCAGCGGCA 873  
 RESULT 11  
 SGPROTG  
 LOCUS 2436 bp DNA linear BCT 21-MAR-1995  
 DEFINITION Streptococcus G148 gene for protein G.  
 ACCESSION X06173 M28331  
 VERSION X06173.1 GI:47084  
 KEYWORDS G protein; IgG receptor.  
 SOURCE Streptococcus sp. G148.  
 ORGANISM Streptococcus sp. G148.  
 Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 Streptococcus.  
 REFERENCE 1 (bases 1 to 2436)  
 AUTHORS Olsson,A., Eliasson,M., Guss,B., Nilsson,B., Hellman,U.,  
 Lindberg,M. and Uhlen,M.  
 TITLE Structure and evolution of the repetitive gene encoding  
 streptococcal protein G  
 JOURNAL Eur. J. Biochem. 168 (2), 319-324 (1987)  
 MEDLINE 88029445  
 FEATURES Location/Qualifiers  
 source  
 1..2436  
 /organism="Streptococcus sp. G148"  
 /strain="G148"  
 /db\_xref="taxon:1324"  
 /clone="pSPG1 and pSPG8"  
 335..340  
 /note="pot. ribosome binding site"  
 346..444  
 346..2127  
 /note="G preprotein (AA -33 to 560)"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="CAA29540.1"  
 /db\_xref="GI:47085"  
 /db\_xref="SWISS-PROT:P19909"  
 /translation="MEKEKKVYFLRKSAFGLASVSAAFVLVGSTVFVAVDSPIEDTPII  
 RRGELTNLLGNSETTLALRNESATADITAAAVADTVAAANAGAAWAAAAAD  
 ALAKAKADALKEFNKYGVSDYKNNLNNAKTVEGVKDLQAQVVEAKKAKISETDGL  
 SDFLSQTPAEDTVKSIELAEAKVLANRELDYGVSDYKNNLNNAKTVEGVKDLQAQ  
 VVEAKKAKISETDGLSDFLSQTPAEDTVKSIELAEAKVLANRELDYGVSDYKNN  
 LNNNAKTVEGVKALIDEILALPDYTKLILNGKTLKGETTTEAVDAATBKVFQY  
 ANDNGVDEWYDDATKTFTVTEKPEVIDASELTPAVTVKLVINGKTLKGETTTEAV  
 DAATAEKVFEQVANDNGVDEWYDDATKTFTVTEKPEVIDASELTPAVTVKLVING  
 KTLKGETTTEAVDAATAEKVFEQVANDNGVDEWYDDATKTFTVTEKPEVIDASELTP  
 EPEKPEASIPVLPVLPATPIAKKDDTKKEDAKKPEAKKEDAKKAEATLPTTGGGS  
 NPFTAAALAVMAGAGALAVASKRED"  
 445..2124  
 /product="mature G protein (AA 1 - 560)"  
 694..765  
 /note="repetitious region 1"  
 766..918  
 /note="repetitious region 2"  
 919..990  
 /note="repetitious region 1"  
 991..1143  
 /note="repetitious region 2"  
 1144..1215  
 /note="repetitious region 1"  
 mat\_peptide  
 repeat\_region  
 repeat\_region  
 repeat\_region  
 repeat\_region  
 repeat\_region

```
/protein_id="AAA03664.1"
/db_xref="GI:153823"
/translation="MEKEKKVYFLRKSAFLASVSAFLVGVSTFVAVDSPIEDTPII
RNGELTNLLGNSETTLALRNEESATADLTAAAVADTVAAAAENAGAAWAAAAAD
ALAKAKADALKEFNKYGVSDYYKNNLNNAKTVEGKIDLAQOVVESAKKARISEATDGL
SDFLSQTPAEDTVKSIELAEAKVLANRELDKYGSDYHKNNLNNAKTVEGKIDLAQ
ILAAALPTDYYKLILNGKTGKGETTTEAVDAATAEKVFKQYANDNGVDGENTYDDATK
TFTVTEKPEVIDASELTPAVTYYKLIVNGKTLKGETTTKAVDAETAEKAFKQYANDNG
VDGVMYTDATKTFTVTMTVEVPCDAPTEPEKPEASIPVPLTPATPIAKDDAKDD
TKKEDAKPEAKKDDAKKAETLPTTGSGNPFPTAALAVMAGALAVASKRKE"
mat_peptide
BASE COUNT 706 a 323 c 398 g 523 t
ORIGIN 1 bp upstream of HindIII site.
```

```
Alignment Scores:
Pred. No.: 3,04e-10 Length: 1950
Score: 155.00 Matches: 37
Percent Similarity: 62.12% Conservative: 4
Best Local Similarity: 56.06% Mismatches: 25
Query Match: 19.95% Indels: 0
DB: 1 Gaps: 0
```

US-09-847-539A-6 (1-159) x STRSPGIGP (1-1950)

```
Qy 1 ValAspSerProileGluGlnProArgileileProAsnGlyGlyThrLeuThrAsnLeu 20
|||||
Db 677 GTTGATTCACCANTCGAGATACCCCAATATTCTGTAATGGTGGCAATTAATCTT 736
|||||
Qy 21 LeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaileAspGluLeu 40
|||||
Db 737 CTGGGAATTCAGACACACACTGGCTTTCGCTAATGAAGAGAGTGCTACAGCTGATTG 796
|||||
Qy 41 LysLysGlnAlaileGluAspLysGluAlaThrThrAlaileGluAlaSerSerAsp 60
|||||
Db 797 ACAGCAGCAGCGGTAGCCGATCTGTGGCAGCAGCGGAGCTGAAATGCTGGGCGACGA 856
|||||
Qy 61 AlaLeuGluAlaLeuAla 66
|||||
Db 857 GCTTGGGAAGCAGCGGCA 874
|||||
```

```
RESULT 8
LOCUS I09115 1950 bp linear PAT 02-DEC-1994
DEFINITION Sequence 23 from Patent WO 8810306.
ACCESSION I09115
VERSION I09115.1 GI:588187
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1950)
AUTHORS Fahnstock,S.R.
TITLE CLONED PROTEIN G VARIANT GENES AND THE PROTEIN G VARIANTS EXPRESSED THEREFROM
JOURNAL Patent: WO 8810306-A 23 29-DEC-1988;
FEATURES Location/Qualifiers
source 1..1950
/organism="unknown"
BASE COUNT 706 a 323 c 398 g 523 t
ORIGIN
```

```
Alignment Scores:
Pred. No.: 3,04e-10 Length: 1950
Score: 155.00 Matches: 37
Percent Similarity: 62.12% Conservative: 4
Best Local Similarity: 56.06% Mismatches: 25
Query Match: 19.95% Indels: 0
DB: 6 Gaps: 0
```

US-09-847-539A-6 (1-159) x I09115 (1-1950)

```
Qy 1 ValAspSerProileGluGlnProArgileileProAsnGlyGlyThrLeuThrAsnLeu 20
```

```
|||||
Db 677 GTTGATTCACCANTCGAGATACCCCAATATTCTGTAATGGTGGCAATTAATCTT 736
|||||
Qy 21 LeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaileAspGluLeu 40
|||||
Db 737 CTGGGAATTCAGACACACACTGGCTTTCGCTAATGAAGAGAGTGCTACAGCTGATTG 796
|||||
Qy 41 LysLysGlnAlaileGluAspLysGluAlaThrThrAlaileGluAlaSerSerAsp 60
|||||
Db 797 ACAGCAGCAGCGGTAGCCGATCTGTGGCAGCAGCGGAGCTGAAATGCTGGGCGACGA 856
|||||
Qy 61 AlaLeuGluAlaLeuAla 66
|||||
Db 857 GCTTGGGAAGCAGCGGCA 874
|||||
RESULT 9
LOCUS SGSPG 2384 bp DNA linear BCT 09-OCT-1997
DEFINITION Streptococcus sp. spg gene for protein G.
ACCESSION Y00428
VERSION Y00428.1 GI:47100
KEYWORDS G protein; IgG binding protein; spg gene.
SOURCE Streptococcus sp. GX7805.
ORGANISM Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae; Streptococcus; Fahnstock,S.R.
REFERENCE 1 (bases 1 to 2384)
AUTHORS Direct Submission
TITLE Submitted (12-OCT-1987) Fahnstock S.R.; Genex Corp., 16020 Industrial Dr., Gaithersburg, MD 20877, USA
REFERENCE 2 (bases 1 to 2384)
AUTHORS Filipula,D., Alexander,P. and Fahnstock,S.R.
TITLE Nucleotide sequence of the protein G gene from Streptococcus GX7805, and comparison to previously reported sequences
JOURNAL Nucleic Acids Res. 15 (17), 7210 (1987)
FEATURES MEDLINE 88015586
Location/Qualifiers
source 1..2384
/organism="Streptococcus sp. GX7805"
/strain="GX7805"
/db_xref="taxon:1325"
577..2358
/gene="spg"
577..675
/gene="spg"
577..2358
/gene="spg"
/codon_start=1
/transl_table=11
/product="protein G precursor"
/protein_id="CAA68489.1"
/db_xref="GI:47101"
/transl_table="MEKEKKVYFLRKSAFLASVSAFLVGVSTFVAVDSPIEDTPII
RNGELTNLLGNSETTLALRNEESATADLTAAAVADTVAAAAENAGAAWAAAAAD
ALAKAKADALKEFNKYGVSDYYKNNLNNAKTVEGKIDLAQOVVESAKKARISEATDGL
SDFLSQTPAEDTVKSIELAEAKVLANRELDKYGSDYHKNNLNNAKTVEGKIDLAQ
VSESAKKARISEATDGLSDFLSQTPAEDTVKSIELAEAKVLANRELDKYGSDYKNN
LNNAKTVEGKIDLAQVSDYYKNNLNNAKTVEGKIDLAQVSESAKKARISEATDGL
ANDNGVDGENTYDDATKTFTVTEKPEVIDASELTPAVTYYKLIVNGKTLKGETTTEAV
DAATAEKAFKQYANDNGVDGENTYDDATKTFTVTEKPEVIDASELTPAVTYYKLIVNG
KTLKGETTTKAVDAETAEKAFKQYANDNGVDGENTYDDATKTFTVTEKPEVIDASELTPAVTYYKLIVNG
EPEKPEASIPVPLTPATPIAKDDAKKDDTKKEDAKPEAKKDDAKKAETLPTTGEGS
NPFTAALAVMAGALAVASKRKE"
676..2355
/gene="spg"
/product="protein G"
886..996
/feature="A1 repeated sequence"
986..1435
/feature="225 bp duplication"
/rpt_unit=1211..1435
mat_peptide
repeat_region
repeat_region
```





AUTHORS Rasmussen,M., Muller,H.P. and Bjorck,L.  
 TITLE Protein GRAB of streptococcus pyogenes regulates proteolysis at the bacterial surface by binding alpha2-macroglobulin  
 JOURNAL J. Biol. Chem. 274 (22), 15336-15344 (1999)  
 MEDLINE 99269061  
 PUBMED 10336419  
 REFERENCE 2 (bases 1 to 469)  
 AUTHORS Rasmussen,M., Muller,H.P. and Bjorck,L.  
 TITLE Direct Submission  
 JOURNAL Submitted (28-JAN-1999) Cell and Molecular Biology, Molecular Pathogenesis, Soelvegatan 39, Lund 221 00, Sweden  
 FEATURES  
 source  
 Location/Qualifiers  
 1..469  
 /organism="Streptococcus pyogenes"  
 /strain="AP1"  
 /db\_xref="taxon:1314"  
 1..>469  
 /product="GRAB"  
 /note="Region: alpha2-macroglobulin-binding A domain"  
 <1..>469  
 /note="grab"  
 /codon\_start=1  
 /transl\_table=11  
 /product="GRAB precursor"  
 /protein\_id="AAD26339.1"  
 /db\_xref="GI:4589081"  
 /translation="VDSPIEQRIIPNGTLLNLGNAPKALRNEERAIDELKKQA  
 IEDKEATTAEAASSDALEALADQTDALQSEAAVVKADNAASDALEALADQTDALQ  
 SEAAVQSDNAAGDALEALADQTDALQSEAAVVKADNAASDALEALADQTDALQSE  
 ASVKADNAASDALEALADQTDALQSEAAVVKADNAASDALEALADQTDALQSEAE  
 VVQSDNAASDAWEKAATPIALDVKRTKDKPVVKKEERQNVNTLPTTGE"  
 175..342  
 misc\_feature  
 /note="Region: repeat motif"  
 misc\_feature 343..469  
 /note="Region: cell-wall attachment"  
 BASE COUNT 178 a 90 c 100 g 101 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 7,16e-88 Length: 469  
 Score: 754.00 Matches: 155  
 Percent Similarity: 99.36% Conservativity: 0  
 Best Local Similarity: 99.36% Mismatches: 1  
 Query Match: 97.04% Indels: 0  
 DB: 1 Gaps: 0  
 US-09-847-539a-6 (1-159) x AF124400 (1-469)  
 QY 1 ValAspSerProIleGluGlnProArgIleIleProAsnGlyGlyThrLeuThrAsnLeu 20  
 |||||  
 DB 1 GTGGACTCACTATCGAACAGCCTAGCAATTATTCCAAATGGCGGAACCTTAATTAATCTT 60  
 |||||  
 QY 21 LeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaIleAspGluLeu 40  
 |||||  
 DB 61 CTGGCAATGCTCCAGAAACCTGGCATTTACGATTGAAGAAGAGCCATTGATGAATTA 120  
 |||||  
 QY 41 LysLysGlnAlaIleGluAspLysGluAlaThrAlaIleGluAlaAlaSerSerAsp 60  
 |||||  
 DB 121 AAAAACACAGCTATTGAGGATTAGGAGCTAGCACAGCTATTAGACCAACCAAGTTCCAGAT 180  
 |||||  
 QY 61 AlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVal 80  
 |||||  
 DB 181 GCCTTAGAAGCATTAGCGGATCAACAGACGCTTTACATCATCAGAACGCTCGGGTTGTT 240  
 |||||  
 QY 81 LysAlaAspAsnAlaAlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeu 100  
 |||||  
 DB 241 AAGCGGATTAACGCTGCTAGTGACGCTTTAGAGCATTTGCGGATCAACAGACGCTTTA 300  
 |||||  
 QY 101 GlnSerGluGluAlaGluValValGlnSerAspAsnAlaAlaSerAspAlaTrpGluLys 120  
 |||||  
 DB 301 CAATCAGAGAGAGCTCAAGTAGTTCATTCAGATAACGCTGCTAGTGACGCTCGGGAAAAA 360  
 |||||  
 QY 121 AlaAlaThrProIleAlaLeuAspValLysLysThrLysAspThrLysProValValLys 140  
 |||||

Db 361 GCAGCAACTCCAATCGCTTTAGATGTTAGAAAACATAAGATACAAAACCTGTAGTTAAA 420  
 QY 141 LysGluGluArgGlnAsnValAsnThrLeuProThrThrGlyGluGlu 156  
 |||||  
 Db 421 AAAGAAGAAGACAAACGTTAATACCTTCCTACAACTGGTGAGAG 468  
 RESULT 5  
 AF124403  
 LOCUS Streptococcus pyogenes strain AP49 GRAB precursor, gene, partial  
 DEFINITION cds.  
 ACCESSION AF124403  
 VERSION AF124403.1 GI:4589086  
 KEYWORDS Streptococcus pyogenes.  
 SOURCE Streptococcus pyogenes  
 ORGANISM Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 Streptococcus.  
 REFERENCE 1 (bases 1 to 804)  
 AUTHORS Rasmussen,M., Muller,H.P. and Bjorck,L.  
 TITLE Protein GRAB of streptococcus pyogenes regulates proteolysis at the bacterial surface by binding alpha2-macroglobulin  
 JOURNAL J. Biol. Chem. 274 (22), 15336-15344 (1999)  
 MEDLINE 99269061  
 PUBMED 10336419  
 REFERENCE 2 (bases 1 to 804)  
 AUTHORS Rasmussen,M., Muller,H.P. and Bjorck,L.  
 TITLE Direct Submission  
 JOURNAL Submitted (28-JAN-1999) Cell and Molecular Biology, Molecular Pathogenesis, Soelvegatan 39, Lund 221 00, Sweden  
 FEATURES  
 Location/Qualifiers  
 source  
 1..804  
 /organism="Streptococcus pyogenes"  
 /strain="AP49"  
 /db\_xref="taxon:1314"  
 1..>804  
 /product="GRAB"  
 1..174  
 /note="Region: alpha2-macroglobulin-binding A domain"  
 <1..>804  
 /note="grab"  
 /codon\_start=1  
 /transl\_table=11  
 /product="GRAB precursor"  
 /protein\_id="AAD26342.1"  
 /db\_xref="GI:4589087"  
 /translation="VDSPIEQRIIPNGTLLNLGNAPKALRNEERAIDELKKQA  
 IEDKEATTAEAASSDALEALADQTDALQSEAAVVKADNAASDALEALADQTDALQSE  
 EAAVQSDNAAGDALEALADQTDALQSEAAVVKADNAASDALEALADQTDALQSEAE  
 ASVKADNAASDALEALADQTDALQSEAAVVKADNAASDALEALADQTDALQSEAE  
 VVQSDNAASDAWEKAATPIALDVKRTKDKPVVKKEERQNVNTLPTTGE"  
 175..678  
 misc\_feature  
 /note="Region: repeat motif"  
 misc\_feature 679..804  
 /note="Region: cell-wall attachment"  
 BASE COUNT 284 a 161 c 185 g 174 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 1.31e-59 Length: 804  
 Score: 536.00 Matches: 113  
 Percent Similarity: 94.26% Conservativity: 2  
 Best Local Similarity: 92.62% Mismatches: 7  
 Query Match: 68.98% Indels: 0  
 DB: 1 Gaps: 0  
 US-09-847-539a-6 (1-159) x AF124403 (1-804)  
 QY 1 ValAspSerProIleGluGlnProArgIleIleProAsnGlyGlyThrLeuThrAsnLeu 20  
 |||||  
 DB 1 GTGGACTCACTATCGAACAGCCTCGAATTTATTCAAAATGGCGGAACCTTAATTAATCTT 60  
 |||||  
 QY 21 LeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaIleAspGluLeu 40

```
Pred. No.: 2,15e-89 Length: 10029
Score: 777.00 Matches: 159
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 1 Gaps: 0

US-09-847-539A-6 (1-159) x AE006573 (1-10029)

Qy 1 ValAspSerProIleGluGlnProArgIleIleProAsnGlyGlyThrLeuThrAsnLeu 20
Db 6482 GTTGACTACCTTATCGACAGCCTCGAATATTCCAAATGGCGAACCTTAACATACTT 6423

Qy 21 LeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaIleAspGluLeu 40
Db 6422 CTTGCAATGCTCCAGAAACCTGCGCATTTACGTAATGAAGAAAGAGCCATGATCAATTA 6363

Qy 41 LysLysGlnAlaIleGluAspLysGluAlaThrAlaIleGluAlaAlaSerSerAsp 60
Db 6362 AAAAACAAGCTATTGAGGATAAAGAGCTAGCAGAGCTATAGAAGCAGCAGCTTTA 6303

Qy 61 AlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaVal 80
Db 6302 GCCTTAGAAGCAATTCGGATCAACAGACGCTTTACAAATCAGAAGAGCTCGGTTGTT 6243

Qy 81 LysAlaAspAsnAlaAlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeu 100
Db 6242 AAAGCGGATAACGCTGCTAGTCACGCTTAGAAGCATTGGCGGATCAACAGCCTTTA 6183

Qy 101 GlnSerGluGluAlaGluValValGlnSerAspAsnAlaAlaSerAspAlaTrpGluLys 120
Db 6182 CAATCAGAAGAAGCTGAAGTAGTTCAATCAGATAACGCTGCTAGTCAGCCTGGGAAAA 6123

Qy 121 AlaAlaThrProIleAlaLeuAspValLysLysThrLysAspThrLysProValValLys 140
Db 6122 GCAGCACTCCAACTGCTTAGATGTTAGAAAACCTAAAGATACAAAACCTGTAGTTAA 6063

Qy 141 LysGluGluArgGlnAsnValAsnThrLeuProThrThrGlyGluGluSerAsnPro 159
Db 6062 AAAGAAGAAAGACAAACGTTAATACCTTCCTACAACTGGTGAAGAGTCTAACCCA 6006

RESULT 3
AF124401 AF124401 468 bp DNA linear BCT 14-AUG-2000
LOCUS Streptococcus pyogenes strain KTL3 GRAB precursor, gene, partial
DEFINITION cds.
ACCESSION AF124401
VERSION AF124401.1 GI:4589082
KEYWORDS Streptococcus pyogenes.
SOURCE Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
ORGANISM Streptococcus.

REFERENCE
1 (bases 1 to 468)
Rasmussen,M., Muller,H.P. and Bjorck,L.
TITLE Protein GRAB of streptococcus pyogenes regulates proteolysis at the
bacterial surface by binding alpha2-macroglobulin
J. Biol. Chem. 274 (22), 15336-15344 (1999)
MEDLINE 99269061
PUBMED 10336419

REFERENCE
2 (bases 1 to 468)
Rasmussen,M., Muller,H.P. and Bjorck,L.
TITLE Direct Submission
JOURNAL Submitted (28-JAN-1999) Cell and Molecular Biology, Molecular
Pathogenesis, Soelvegatan 39, Lund 221 00, Sweden
FEATURES
source Location/Qualifiers
1..468
/organism="Streptococcus pyogenes"
/strain="KTL3"
/db_xref="taxon:1314"
mat_peptide 1..>468
misc_feature 1..174

Pred. No.: 1,21e-88 Length: 468
Score: 760.00 Matches: 156
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 97.81% Indels: 0
DB: 1 Gaps: 0

US-09-847-539A-6 (1-159) x AF124401 (1-468)

Qy 1 ValAspSerProIleGluGlnProArgIleIleProAsnGlyGlyThrLeuThrAsnLeu 20
Db 1 GTGGACTCACCTATCGAACAGCCTCGAATATTCCAAATGGCGAACCTTAACATACTT 60

Qy 21 LeuGlyAsnAlaProGluLysLeuAlaLeuArgAsnGluGluArgAlaIleAspGluLeu 40
Db 61 CTTGCAATGCTCCAGAAACCTGCGCATTTACGTAATGAAGAAAGAGCCATGATCAATTA 120

Qy 41 LysLysGlnAlaIleGluAspLysGluAlaThrAlaIleGluAlaAlaSerSerAsp 60
Db 121 AAAAACAAGCTATTGAGGATAAAGAGCTAGCAGAGCTATAGAAGCAGCAGCTTCAGAT 180

Qy 61 AlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaVal 80
Db 181 GCCTTAGAAGCAATTCGGATCAACAGCCTTTACAAATCAGAAGAGCTCGGTTGTT 240

Qy 81 LysAlaAspAsnAlaAlaSerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeu 100
Db 241 AAAGCGGATAACGCTGCTAGTCACGCTTAGAAGCATTGGCGGATCAACAGCCTTTA 300

Qy 101 GlnSerGluGluAlaGluValValGlnSerAspAsnAlaAlaSerAspAlaTrpGluLys 120
Db 301 CAATCAGAAGAAGCTGAAGTAGTTCAATCAGATAACGCTGCTAGTCAGCCTGGGAAAA 360

Qy 121 AlaAlaThrProIleAlaLeuAspValLysLysThrLysAspThrLysProValValLys 140
Db 361 GCAGCACTCCAACTGCTTAGATGTTAGAAAACCTAAAGATACAAAACCTGTAGTTAA 420

Qy 141 LysGluGluArgGlnAsnValAsnThrLeuProThrThrGlyGluGlu 156
Db 421 AAAGAAGAAAGACAAACGTTAATACCTTCCTACAACTGGTGAAGAG 468

RESULT 4
AF124400 AF124400 469 bp DNA linear BCT 14-AUG-2000
LOCUS Streptococcus pyogenes strain AP1 GRAB precursor, gene, partial
DEFINITION cds.
ACCESSION AF124400
VERSION AF124400.1 GI:4589080
KEYWORDS Streptococcus pyogenes.
SOURCE Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
ORGANISM Streptococcus.

REFERENCE
1 (bases 1 to 469)
```

```

CDS
complement(619..1911)
/genes="aroA"
/EC_number="2.5.1.19"
/notes="Best Blastp hit = sp|P43905|AROA_LACLA
3-PHOSPHOSHIMIMATE 1-CARBOXYVINYLTRANSFERASE
(5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE) (EPSP
SYNTHASE) (EPSP) >gi|1075724|pir|J52580
3-phosphoshikimate 1-carboxyvinyltransferase (EC 2.5.1.19)
- Lactococcus lactis >gi|683583|emb|CAA55180.1| (X78413)
5-enolpyruvylshikimate-3-phosphate synthase [Lactococcus
lactis]
/codon_start=1
/transl_table=11
/product="putative 3-phosphoshikimate
1-carboxyvinyltransferase"
/protein_id="AAK34180.1"
/db_xref="GI:13622461"
/translation="MKRMKLTWAGPLOGTIQVQDKSISHRAVILGAVAKGETRVKG
LKGEDVLSITQAFNLGVRIEKKDQVIEGQFQGLNAPCOTLNGNSGTSKRLIA
GLLAGQFVYKMTGDESLSKRPMDRIVYPLKQMGVEISGETDRQFPPLQIQGNRLQP
ITVLPISSAQVKSAILLAALQAKGTQVVEKEITRNHTEEMIQOFGGRLIVGKRIT
LVGPQOLTAQEIIVPGDISAAPMLVAGLIIPGSELLKNVGNVPTTGTGLEVERKG
AQIYVEDMKKEQVTSIRVYVSNMKGTTISGGULPRILDELPIIALLAQAQCTTICK
DAQLRVKTRDIQVVTIDILNSMGANIKATADGMI1KGPTVLYGANTSTYGDHRIQHM
TATAALLVQGVHLDKEAINTSYPTFFKDLERLCHD"
/complement(2013..2978)
/genes="SPY1353"
complement(2013..2978)
/genes="SPY1353"
/notes="Best Blastp hit = dbj|BAB07074.1| (AP001518)
unknown conserved protein [Bacillus halodurans]
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK34181.1"
/db_xref="GI:13622462"
/translation="MAEKKVDFKVLKSWQYEQVPMRHLQSAEMDLAIAVAYLIL
TAPPLVIAIAPPYINIDIALRLMKONLPKDIKDPASAIENIFSKPSGVLGVA
TLTGLWMSRLSTLOKAIKAYGASQHRDFFIGHVLGLTSLILFLFAFLAIFIF
SKAAIOLVKHLSNITIFILLQIPTVLIIFVGLMLLFLFLNPKIKKRIYLP
GTLFTSFVTFISLNVGNVYVNERVDKMGVSQIFIMLWIFLAKILILGAIF
NATYQMSIGLKGEGSDGMIALILKTLGNSDLSPSQSIEDSHTD"
/complement(2980..3840)
/genes="map"
/notes="SPY1354"
complement(2980..3840)
/genes="map"
/EC_number="3.4.11.18"
/notes="Best Blastp hit = sp|Q929J4|AMPM_BACHD METHIONINE
AMINOPEPTIDASE (MAP) (PEPTIDASE M)
>gi|11264635|pir|J74405 methionyl aminopeptidase (EC
3.4.11.18) map [imported] - Bacillus halodurans
>gi|4512426|dbj|JBA75293.1| (AB017508) map homologue
(identity of 81% to B. subtilis ) [Bacillus halodurans]
>gi|10172768|dbj|BAB03875.1| (AP001507) methionine
aminopeptidase [Bacillus halodurans]
/codon_start=1
/transl_table=11
/product="putative methionine aminopeptidase"
/protein_id="AAK34182.1"
/db_xref="GI:13622463"
/translation="MTLKSAREIEAMDAGDFLAGIHGLRDLIRKPGVDMWEVAYV
RRCKEDNVLPIQIGVDGHMDPYATCCGLNDEVAHAFPHRYILKEGLLKVDVYLS
EPLDKSIVDYAALDFENYEMKWTGSGYGGGLADSCWATVAGTPSDIEKQMLDKTEA
MYRIGKAVIGNRINDIGAAVYAESFGYGVVRLDVGHVGTTHHEEPMPVNYGTAG
RGLRKEGMVLVTEPMINTGTWEIDTIDIKTGAHKTLDGGLSCQYEHQFVITKDGPIV
LTSQGEERTY"
complement(3856..5139)
/genes="SPY1355"
complement(3856..5139)
/genes="SPY1355"
/notes="Best Blastp hit = pir|A69998 hypothetical protein
ytoi - Bacillus subtilis >gi|2293259|gb|AAC00336.1|
(AF008220) Ytoi [Bacillus subtilis]
>gi|2635392|emb|CAB14887.1| (Z99118) ytoi [Bacillus
subtilis]
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK34183.1"
/db_xref="GI:13622464"
/translation="MSKHODILQYLEKLAIGKVSVRSTSNHLKVSQDGTAYRAIKAE
NEGIVETKPSGTVRIEKKGRVIRDTLSEIARHNSIDSEVLAGHGFHPSRFSIGA
MTQONRRYLVKGLLIVGDRITQILLALENNAILVTGGFPVSKRVIEHANNQRIPV
MVTHYDTFVATMINHALSNIRIKTDLKTVEQVMIPTDYGLCEDSSVEEFTNLIKK
TRQVREPVDYDKRKVICVVSMDRVQDLPTTKLTVMKSNPITARPNTSIANISQKMI
FEDLNMLPVTDENNLLGMITRRAEMENLPNHPNPTYSQIILSNLETVQYOVV
VEPTMIDSNAGNSNGVISELKEISIRALTQKHKNTIIIEQNMVYFLHAQIEDELKI
YPLIITENRRSTIDIEIFVDDQVIAKAITTKIN"
/complement(5148..5690)
/genes="SPY1356"
complement(5148..5690)
/genes="SPY1356"
/notes="Best Blastp hit = emb|CAA03913.1| (AJ000084)
putative acetyl transferase [Proteus mirabilis]
/codon_start=1
/transl_table=11
/product="putative acetyl transferase"
/protein_id="AAK34184.1"
/db_xref="GI:13622465"
/translation="MDIWTKLAVFAFETPKVILRPFRYEDHWDFTSMVNDTKNLYYV
FPEQTKAASDYLLVHSFKFPLGOWAIEDKATHQVIGSIRIEHYDAKTADIGYFL
NFAFGQGITVEVYIKLVYLSFHEFGTLRIITHLENKASQKVAKKAGCOLKTCFPG
SDRNTHKICIKMYQLTNDK"
/complement(5928..6581)
/genes="grab"
/notes="SPY1357"
complement(5928..6581)
/genes="grab"
/notes="Best Blastp hit = gb|AAD26340.1|AF124401.1
(AF124401) GRAB precursor [Streptococcus pyogenes]"
/codon_start=1
/transl_table=11
/product="protein GRAB (protein G-related alpha 2W-binding
protein)"
/protein_id="AAK34185.1"
/db_xref="GI:13622466"
/translation="MKKEIKVCFLLRSRAGFLVAVSASVLVSTVSASVSPIDQPII
PNGGTLNLLGNAPKRLALRNEERAIDELKQAIEDKEATTATEAASSDALEALADOT
DALQSEEAIVVVRADNAADALEALADOTDALQSEEAIVVQSDNAASDAWEKAATPIAL
DVKKTKDTPVVKKEERQNVNTLPTTGEESNPFTTAAALAIMVSTGLVLYVSSCKEN"
/complement(6939..8198)
/genes="murZ"
/notes="SPY1358"
complement(6939..8198)
/genes="murZ"
/notes="Best Blastp hit = sp|P19670|MUA2_BACSU
UDP-N-ACETYLGLUCOSAMINE 1-CARBOXYVINYLTRANSFERASE 2
(ENOYL PYRUVATE TRANSFERASE) (UDP-N-ACETYLGLUCOSAMINE
ENOYL PYRUVYL TRANSFERASE) (EPT) >gi|2127262|pir|J32354
UDP-N-acetylglucosamine 1-carboxyvinyltransferase (EC
2.5.1.7) murZ - Bacillus subtilis
>gi|853767|emb|CAA89875.1| (Z49782)
UDP-N-acetylglucosamine 1-carboxyvinyltransferase
[Bacillus subtilis] >gi|2636235|emb|CAB15727.1| (Z99122)
UDP-N-acetylglucosamine 1-carboxyvinyltransferase
[Bacillus subtilis] >gi|2636247|emb|CAB15738.1| (Z99123)
UDP-N-acetylglucosamine 1-carboxyvinyltransferase
[Bacillus subtilis]
/codon_start=1
/transl_table=11
/product="putative UDP-N-acetylglucosamine
1-carboxyvinyltransferase"
/protein_id="AAK34186.1"

```

Alignment Scores:



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 13, 2002, 02:12:06 ; Search time 1551.74 Seconds  
(without alignments)  
2144.253 Million cell updates/sec

Title: US-09-847-539A-6

Perfect score: 777

Sequence: 1 VDSPIEQPIIPNGGTLNL.....KKEERQNVNTLPTTGESNP 159

Scoring table:

BLOSUM62  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DRV=vlh  
-Q/cgn2\_1/USPTO.spool/US09847539/runat\_10102002\_092548\_3423/app\_query.fasta\_1.526  
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=40 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09847539 -CGN\_1\_1182\_@runat\_10102002\_092548\_3423 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

GenEmbl: \*  
1: gb\_ba.\*  
2: gb\_htg.\*  
3: gb\_in.\*  
4: gb\_om.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vl.\*  
15: em\_ba.\*  
16: em\_fun.\*  
17: em\_hum.\*  
18: em\_in.\*  
19: em\_mu.\*  
20: em\_om.\*  
21: em\_or.\*  
22: em\_ov.\*  
23: em\_pat.\*  
24: em\_ph.\*  
25: em\_pi.\*  
26: em\_ro.\*  
27: em\_sts.\*  
28: em\_un.\*

29: em\_vi.\*  
30: em\_htg\_hum.\*  
31: em\_htg\_inv.\*  
32: em\_htg\_other.\*  
33: em\_htgo\_inv.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	777	100.0	832	1	AF124399 Streptococcus
2	777	100.0	10029	1	AE006573 Streptococcus
3	760	97.8	468	1	AF124401 Streptococcus
4	754	97.0	469	1	AF124400 Streptococcus
5	536	69.0	804	1	AF124403 Streptococcus
6	534	68.7	717	1	AF124402 Streptococcus
7	155	19.9	1950	1	STRSPGIGP
8	155	19.9	1950	6	I09115
9	155	19.9	2384	1	SGSPG
10	155	19.9	2384	6	I08537
11	155	19.9	2456	1	SGPROTG
12	153	19.7	1950	6	I08536
13	142	18.3	1576	1	SGI48PG
14	126	16.2	969	1	STRABP
15	113	14.5	167227	9	AC020602
16	105	13.5	798	6	I09107
17	104	13.4	3222	14	HPU23857
18	104	13.4	106516	1	STYPPHMC2
19	99	12.7	48502	7	LAMCG
20	96	12.4	13501	1	AE005258
21	96	12.4	297816	1	AP002553
22	92	11.8	220820	2	AL663060
23	91	11.7	1555	1	STRNAG
24	90	11.6	1469	1	SGI48IGG
25	90	11.6	1469	6	AL2446
26	90	11.6	2022	3	AY060997
27	90	11.6	6484	3	AF335500
28	90	11.6	10432	1	AE005333
29	90	11.6	54784	2	AC017732
30	90	11.6	140973	3	AC008140
31	90	11.6	179669	3	AC008089
32	90	11.6	222605	1	AP002555
33	90	11.6	224795	3	AE003688
34	90	11.6	269726	2	AC020883
35	89	11.5	2437	1	SAG277292
36	89	11.5	9331	1	AE006458
37	89	11.5	20777	2	LMFLCHR34_17
38	89	11.5	110000	2	LMFLCHR34_16
39	89	11.5	291136	1	AP002556
40	88	11.3	848	11	CNS0618W
41	88	11.3	3720	14	S76368
42	88	11.3	43658	14	HSV3PRGEN
43	88	11.3	112930	14	HSGEND
44	88	11.3	189129	2	AC022682
45	88	11.3	256608	10	AL589699

#### ALIGNMENTS

RESULT 1

AF124399

LOCUS

DEFINITION

complete cds.

ACCESSION

AF124399

VERSION

AF124399.1

KEYWORDS

SOURCE

AF124399 832 bp DNA linear BCT 14-AUG-2000  
Streptococcus pyogenes strain ATCC700294 GRAB precursor, gene,  
complete cds.  
AF124399 GI:4589078  
Streptococcus pyogenes.

XX  
SQ Sequence 200 AA;  
Query Match 10.7%; Score 83; DB 16; Length 200;  
Best Local Similarity 54.8%; Pred. No. 0.043;  
Matches 17; Conservative 2; Mismatches 12; Indels 0; Gaps 0;  
Qy 128 DVKTKDTKPVVYKKEERONVNTLPTTGEESN 158  
| | | | | : | | | | |  
Db 163 DTKKEDAKKPEAKKEDAKKAETLPTTGEESN 193  
Search completed: October 13, 2002, 02:08:14  
Job time : 66.6203 secs







PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140931.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.

PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159283.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.



```
XX (HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YT;
PI WPI; 2001-639362/73.
XX N-PSDB; AAS82434.
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX Claim 20; SEQ ID No 48606; 103pp; English.
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG0010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 1107 AA:
Query Match 11.1%; Score 86; DB 22; Length 1107;
Best Local Similarity 26.8%; Pred. No. 0.15;
Matches 26; Conservative 16; Mismatches 55; Indels 0; Gaps 0;
QY 46 EDKEATTATEAASDALEALADOTDALQSEAAVVRADNAADALEALADOTDALQSEEA 105
DB 647 QSSAAASSASAATSETNAKASETSAESKTAASASSASSASSASSASASAKDEATRQASA 706
QY 106 EVVQSDNAASDAWEKAATPTALDVKKTKDTPVVKKE 142
DB 707 AKGSATTASTKATEAAGSATAAQAQSKSTAESAAATRAE 743
RESULT 35
AAR71929
ID AAR71929 standard; Protein; 664 AA.
XX
AC AAR71929;
XX
XX 22-SEP-1995 (first entry)
XX
DE S. dysgalactiae MIG.
XX
KW MIG: fast alpha-2-macroglobulin binding protein; FAM;
KW plasma proteinase-inhibitor binding protein.
XX
OS Streptococcus dysgalactiae strain SC1.
XX
FH Key Location/Qualifiers
FT Peptide 1..30
FT /label= sig_peptide
FT 209..278
FT Domain
FT /label= IgG1
FT /note= "IgG binding domain 1"
```

```
FT Domain 279..348
FT /label= IgG2
FT /note= "IgG binding domain 2"
FT 349..418
FT /label= IgG3
FT /note= "IgG binding domain 3"
FT 419..488
FT /label= IgG4
FT /note= "IgG binding domain 4"
FT 489..558
FT /label= IgG5
FT /note= "IgG binding domain 5"
FT 559..638
FT /note= "cell wall spanning region"
FT 639..664
FT /note= "putative wall anchoring motif"
FT 639..664
FT /note= "membrane spanning domain"
XX
XX WO9507296-A.
XX
XX 16-MAR-1995.
XX
XX 06-SEP-1994; 94WO-SE00826.
XX
XX 06-SEP-1993; 93SE-0002855.
XX
XX (GUSS/) GUSS B.
XX (JONS/) JONSSON H.
XX (LIND/) LINDBERG M.
XX (MUEL/) MUELLER H.
XX (RANT/) RANTAMAKI L K.
XX
XX Guss B, Jonsson H, Lindberg M, Mueller H, Rantamaki LK;
XX WPI; 1995-123382/16.
XX N-PSDB; AAQ89196.
XX
XX DNA encoding fast alpha 2-macroglobulin-binding proteins - used
XX to obtain prods. for sepn., detection or quantification or for
XX binding inhibition
XX
XX Disclosure; Page 31; 50pp; English.
XX
XX A phage lambda GEM-11 library of S. dysgalactiae SC1 DNA was analyzed
XX for fast-acting alpha-2-macroglobulin (FAM)- and IgG-binding
XX activity. One clone, which expressed both activities, was
XX analyzed to obtain DNA encoding the FAM-binding protein, MIG
XX (AAR71929). The mig gene is given in AAQ89196.
XX
XX Sequence 664 AA;
Query Match 10.9%; Score 85; DB 16; Length 664;
Best Local Similarity 56.2%; Pred. No. 0.1;
Matches 18; Conservative 2; Mismatches 12; Indels 0; Gaps 0;
QY 128 DVKTKTDKTPVVKKEERQNVNLTPTTGESNP 159
DB 608 DAKKPEAKKEAKKEAKKAATLTPTTGESNP 639
RESULT 36
ABB58655
ID ABB58655 standard; Protein; 489 AA.
XX
AC ABB58655;
XX
XX 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 2757.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
```

```

XX SQ Sequence 413 AA;
Query Match 11.7%; Score 91; DB 16; Length 413;
Best Local Similarity 59.4%; Pred. No. 0.011;
Matches 19; Conservative 1; Mismatches 12; Indels 0; Gaps 0;

Qy 128 DVKTKDTKPVVKKKEERONVNTLPTTGESNP 159
   || ||| : ||||| |||
Db 357 DTKKEDAKKPEAKKEAKAATLPTTGESNP 388

RESULT 32
AAP70468
ID AAP70468 standard; Protein; 480 AA.
XX
AC AAP70468;
XX
DT 21-MAY-1991 (first entry)
XX
DE Sequence of polypeptide possessing IgG-binding activity of
DE protein G from Streptococcus G148.
XX
KW Antibody-binding; IgG; IgA; immunoglobulin.
XX
OS Streptococcus G148.
XX
PN WO8705631-A.
XX
PD 24-SEP-1987.
XX
PF 20-MAR-1987; 87WO-SE00145.
XX
PR 21-MAR-1986; 86SE-0001325.
XX
PA (PHAA ) PHARMACIA AB.
PA (GUSS/) GUSS B M.
XX
PI Guss BM, Lindberg KM, Flock JI, Uhlen CEM;
XX
DR WPI; 1987-277686/39.
DR N-PSDB; AAN70757.
XX
PT New recombinant DNA molecules - for producing proteins with
PT IgG-binding specificity of protein G or proteins A and G
XX
PS Example; Fig 2; 39pp; English.
XX
CC A recombinant DNA molecule containing a nucleotide sequence which
CC codes for a protein or polypeptide having the same IgG specificity as
CC protein G from Streptococcus G148 (AAN70757) is claimed. See, for
CC example, AAN70754, AAN70755 and AAN70756.
XX
SQ Sequence 480 AA;
Query Match 11.6%; Score 90; DB 8; Length 480;
Best Local Similarity 56.2%; Pred. No. 0.017;
Matches 18; Conservative 2; Mismatches 12; Indels 0; Gaps 0;

Qy 128 DVKTKDTKPVVKKKEERONVNTLPTTGESNP 159
   || ||| : ||||| |||
Db 424 DTKKEDAKKPEAKKEAKAATLPTTGESNP 455

RESULT 33
AAV71049
ID AAV71049 standard; peptide; 19 AA.
XX
AC AAV71049;
XX
DT 29-AUG-2000 (first entry)
XX
DE Streptococcus pyogenes GRAB peptide EKL 18.

```

```

XX GRAB; protein G related alpha2M binding protein; vaccine;
KW alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;
KW immune response; Streptococcus pyogenes infection; peptide EKL 18.
XX Streptococcus pyogenes.
OS WO200026240-A2.
XX
PN 11-MAY-2000.
XX
PF 02-NOV-1999; 99WO-GB03631.
XX
PR 02-NOV-1998; 98GB-0023975.
XX
PA (ACTI-) ACTINOVA LTD.
XX
PI Bjorck LH, Rasmussen M;
XX
DR WPI; 2000-365572/31.
XX
XX New alpha2M binding protein for generating a protective immune response
PT to group A streptococcus and purifying the binding protein
PT
XX
PS Example 9; Page 31; 67pp; English.
XX
CC The patent discloses a new family of proteins termed GRAB (protein G
CC related alpha2M binding protein) from Streptococcus pyogenes which have
CC the ability to bind alpha2-macroglobulin (alpha2M) and show homology to
CC protein G of group G Streptococcus. GRAB protein and peptides derived
CC from it are used in vaccine compositions for generating a protective
CC immune response against group A Streptococcus. Antibodies against GRAB
CC are useful for treating Streptococcus pyogenes infections. The protein
CC is also useful for purifying alpha2M from a sample. The present sequence
CC is a GRAB peptide EKL 18 useful in vaccine composition. It was used
CC to produce immunoglobulin G (IgG) antibodies specific for native GRAB
CC protein in sheep. The peptide has a cysteine insert at the C-terminal
CC for attachment to a hetero-bifunctional linker.
XX
SQ Sequence 19 AA;
Query Match 11.1%; Score 86; DB 21; Length 19;
Best Local Similarity 100.0%; Pred. No. 0.0011;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 26 EKALRNEERAIDELKKQ 43
   ||||| ||||| |||||
Db 1 EKALRNEERAIDELKKQ 18

RESULT 34
ABG18247
ID ABG18247 standard; Protein; 1107 AA.
XX
AC ABG18247;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #18238.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.

```





```
PR 19-JUN-1987; 87US-0063959.
PR 20-JUN-1988; 88US-0209236.
PR 19-JUN-1990; 90US-0540169.
PR 21-APR-1992; 92US-0871539.
XX
PA (PHAA ) PHARMACIA LKB BIOTECHNOLOGY AB.
XX
PI Farnestock SR;
XX
DR WPI: 1994-159179/19.
DR N-PSDB; AAQ64648.
XX
PT New recombinant streptococcal protein G variants - useful for
PT antibody detection and purification and for therapy
XX
PS Claim 3: Column 46; 48pp; English.
XX
CC Protein G isolated from Streptococcus has IgG-binding activity which
CC has been localised to the B repeating structure (see AAR53290).
CC The sequence AAR53294 represents a claimed Streptococcal Protein G
CC variant comprising the B domain active site and retaining
CC IgG-binding activity.
XX
SQ Sequence 265 AA;
Query Match 13.5%; Score 105; DB 15; Length 265;
Best Local Similarity 69.7%; Pred. No. 0.00012;
Matches 23; Conservative 1; Mismatches 9; Indels 0; Gaps 0;
QY 4 PIERPRIPIPGGTLTLNLLGNAPEKALRNEERA 36
III III III III III III III III III
DB 34 PIEDTPIIRNGGELTNLLGNSETTLALRNEESA 66
RESULT 26
AAY71051
ID AAY71051 standard; peptide; 20 AA.
XX
AC AAY71051;
XX
DT 29-AUG-2000 (first entry)
XX
DE Streptococcus pyogenes GRAB peptide KKT 19.
XX
KW GRAB; protein G related alpha2M binding protein; vaccine;
KW alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;
KW immune response; Streptococcus pyogenes infection; peptide KKT 19.
XX
OS Streptococcus pyogenes.
XX
PN WO200026240-A2.
XX
PD 11-MAY-2000.
XX
PF 02-NOV-1999; 99WO-GB03631.
XX
PR 02-NOV-1998; 98GB-0023975.
XX
PA (ACTI-) ACTINOVA LTD.
XX
PI Bjorck LH, Rasmussen M;
XX
DR Streptococcus pyogenes.
XX
DR WO200026240-A2.
XX
PD 11-MAY-2000.
XX
PF 02-NOV-1999; 99WO-GB03631.
XX
PR 02-NOV-1998; 98GB-0023975.
XX
PA (ACTI-) ACTINOVA LTD.
XX
PI Bjorck LH, Rasmussen M;
XX
DR WPI: 2000-365572/31.
XX
PT New alpha2M binding protein for generating a protective immune response
PT to group A streptococcus and purifying the binding protein -
XX
PS Example 9; Page 31; 67pp; English.
XX
CC The patent discloses a new family of proteins termed GRAB (protein G
CC related alpha2M binding protein) from Streptococcus pyogenes which have
CC the ability to bind alpha2-macroglobulin (alpha2M) and show homology to
CC protein G of group G Streptococcus. GRAB protein and peptides derived
CC from it are used in vaccine compositions for generating a protective
CC immune response against group A Streptococcus. Antibodies against GRAB
CC are useful for treating Streptococcus pyogenes infections. The protein
CC is also useful for purifying alpha2M from a sample. The present sequence
CC is a GRAB peptide DSP 18 useful in vaccine composition. It was used
CC to produce immunoglobulin G (IgG) antibodies specific for native GRAB
CC protein in sheep. The peptide has a cysteine insert at the C-terminal
CC for attachment to a hetero-bifunctional linker.
XX
SQ Sequence 19 AA;
Query Match 12.4%; Score 96; DB 21; Length 19;
Best Local Similarity 100.0%; Pred. No. 6.3e-05;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
CC from it are used in vaccine compositions for generating a protective
CC immune response against group A Streptococcus. Antibodies against GRAB
CC are useful for treating Streptococcus pyogenes infections. The protein
CC is also useful for purifying alpha2M from a sample. The present sequence
CC is a GRAB peptide KKT 19 useful in vaccine composition. It was used
CC to produce immunoglobulin G (IgG) antibodies specific for native GRAB
CC protein in sheep. The peptide has a cysteine insert at the C-terminal
CC for attachment to a hetero-bifunctional linker.
XX
SQ Sequence 20 AA;
Query Match 12.5%; Score 97; DB 21; Length 20;
Best Local Similarity 100.0%; Pred. No. 5e-05;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 130 KTKDKTKPVVKKKEERQNVN 148
IIIIIIIIIIIIIIIIIIII
DB 1 KTKDKTKPVVKKKEERQNVN 19
RESULT 27
AAY71047
ID AAY71047 standard; peptide; 19 AA.
XX
AC AAY71047;
XX
DT 29-AUG-2000 (first entry)
XX
DE Streptococcus pyogenes GRAB peptide DSP 18.
XX
KW GRAB; protein G related alpha2M binding protein; vaccine;
KW alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;
KW immune response; Streptococcus pyogenes infection; peptide DSP 18.
XX
OS Streptococcus pyogenes.
XX
PN WO200026240-A2.
XX
PD 11-MAY-2000.
XX
PF 02-NOV-1999; 99WO-GB03631.
XX
PR 02-NOV-1998; 98GB-0023975.
XX
PA (ACTI-) ACTINOVA LTD.
XX
PI Bjorck LH, Rasmussen M;
XX
DR WPI: 2000-365572/31.
XX
DR New alpha2M binding protein for generating a protective immune response
DR to group A streptococcus and purifying the binding protein -
XX
PS Example 9; Page 31; 67pp; English.
XX
CC The patent discloses a new family of proteins termed GRAB (protein G
CC related alpha2M binding protein) from Streptococcus pyogenes which have
CC the ability to bind alpha2-macroglobulin (alpha2M) and show homology to
CC protein G of group G Streptococcus. GRAB protein and peptides derived
CC from it are used in vaccine compositions for generating a protective
CC immune response against group A Streptococcus. Antibodies against GRAB
CC are useful for treating Streptococcus pyogenes infections. The protein
CC is also useful for purifying alpha2M from a sample. The present sequence
CC is a GRAB peptide DSP 18 useful in vaccine composition. It was used
CC to produce immunoglobulin G (IgG) antibodies specific for native GRAB
CC protein in sheep. The peptide has a cysteine insert at the C-terminal
CC for attachment to a hetero-bifunctional linker.
XX
SQ Sequence 19 AA;
Query Match 12.4%; Score 96; DB 21; Length 19;
Best Local Similarity 100.0%; Pred. No. 6.3e-05;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```



```
Matches 23; Conservative 1; Mismatches 9; Indels 0; Gaps 0;
QY 4 PTEQPIIPNGGTLTLNGLNAPEKALRNEERA 36
   III I I I I I I I I I I I I I I I I
Db 4 PIEDPIIRNGGELTNLGNSETTLALRNEESA 36

RESULT 23
AAR07004
ID AAR07004 standard; protein; 235 AA.
XX
AC AAR07004;
XX
DT 17-JAN-1991 (first entry)
XX
DE Protein G variant.
XX
KW Immunoglobulin.
XX
OS Streptococcus sp. Lancefield Group G strain.
XX
PN US4956296-A.
XX
PD 11-SEP-1990.
XX
PF 20-JUN-1988; 88US-0209236.
XX
PR 20-JUN-1988; 88US-0209236.
PR 14-FEB-1986; 86US-0829354.
PR 23-APR-1986; 86US-0854887.
PR 17-FEB-1987; 87WO-US00329.
PR 19-JUN-1987; 87US-0063959.
XX
PA (GENE-) GENEX CORP.
XX
PI Farnestock SR;
XX
DR WPI; 1990-297491/39.
DR N-PSDB; AAQ06009.
XX
XX Recombinant Protein G variants - obtd. using a cloned gene
PT encoding Protein G from Streptococcus sp., used for binding
PT immunoglobulin.
XX
PS Disclosure; Column 9-16; 48pp; English.
XX
CC Sequence may be incorporated into a non-pathogenic host eg. E.coli,
CC where they may be expressed at high levels. The proteins have a
CC higher binding efficiency and capacity for immunoglobulin, and may
CC be used for purifying, detecting and isolating antibodies.
XX
SQ Sequence 235 AA;
Query Match 13.5%; Score 105; DB 11; Length 235;
Best Local Similarity 69.7%; Pred. No. 0.0001;
Matches 23; Conservative 1; Mismatches 9; Indels 0; Gaps 0;
QY 4 PTEQPIIPNGGTLTLNGLNAPEKALRNEERA 36
   III I I I I I I I I I I I I I I I I
Db 4 PIEDPIIRNGGELTNLGNSETTLALRNEESA 36

RESULT 24
AAR10011
ID AAR10011 standard; Protein; 265 AA.
XX
AC AAR10011;
XX
DT 13-MAR-1991 (first entry)
XX
DE Type 4 Gx7809 protein G variant.
XX
KW Immunoglobulins; Ig.
```

```
XX
OS Streptococcus sp Gx7809.
XX
PN US4977247-A.
XX
PD 11-DEC-1990.
XX
PF 19-MAY-1989; 89US-0354264.
XX
PR 19-MAY-1989; 89US-0354264.
PR 14-FEB-1986; 86US-0829354.
PR 23-APR-1986; 86US-0854887.
PR 17-FEB-1987; 87WO-US00329.
PR 19-JUN-1987; 87US-0063959.
PR 20-JUN-1988; 88US-0209236.
XX
PA (GENE-) GENEX CORP.
XX
PI Farnestock SR, Lee T, Wroble MH;
XX
DR WPI; 1991-006758/01.
DR Q-PSDB; Q10007.
XX
PT Immobilised protein G variants - used for detection, isolation
PT and purificn. immunoglobulin(s) and immunoglobulin fragments
XX
PS Disclosure; Column 11; 52pp; English.
XX
CC Protein G variant product carries active regions B1 and B2, it may
CC be immobilised and exhibits different binding profiles. The bound
CC protein is useful in purification and detection of Igs and fragments.
XX
SQ Sequence 265 AA;
Query Match 13.5%; Score 105; DB 12; Length 265;
Best Local Similarity 69.7%; Pred. No. 0.00012;
Matches 23; Conservative 1; Mismatches 9; Indels 0; Gaps 0;
QY 4 PTEQPIIPNGGTLTLNGLNAPEKALRNEERA 36
   III I I I I I I I I I I I I I I I I
Db 34 PIEDPIIRNGGELTNLGNSETTLALRNEESA 66

RESULT 25
AAR53294
ID AAR53294 standard; Protein; 265 AA.
XX
AC AAR53294;
XX
DT 06-JAN-1995 (first entry)
XX
DE IgG-binding Streptococcus Protein G variant.
XX
KW Streptococcus Protein G; variant; IgG binding activity;
KW immunoglobulin; Lancefield Group G; bacterial Fc receptor.
XX
OS Streptococcus sp.
XX
FH Key Location/Qualifiers
FT Peptide 1..30
FT Protein /label= secretion_sequence
FT /label= 31..265
FT /label= Protein_G_variant
FT /note= "claimed without the secretion sequence"
XX
PN US5312901-A.
XX
PD 17-MAY-1994.
XX
PF 14-FEB-1986; 86US-0829354.
XX
PR 14-FEB-1986; 86US-0829354.
PR 23-APR-1986; 86US-0854887.
```

XX WO200026240-A2.  
 PN 11-MAY-2000.  
 PD  
 XX  
 XX 02-NOV-1999; 99WO-GB03631.  
 PF  
 XX 02-NOV-1998; 98GB-0023975.  
 PR  
 XX (ACTI-) ACTINOVA LTD.  
 PA  
 XX Bjorck LH, Rasmussen M;  
 PI  
 XX WPI; 2000-365572/31.  
 DR  
 XX New alpha2M binding protein for generating a protective immune response  
 PT to group A streptococcus and purifying the binding protein -  
 XX  
 XX Claim 1; Page 55; 67pp; English.  
 PS  
 XX The patent discloses a new family of proteins termed GRAB (protein G  
 CC related alpha2M binding protein) from Streptococcus pyogenes which have  
 CC the ability to bind alpha2-macroglobulin (alpha2M) and show homology to  
 CC protein G of group G Streptococcus. GRAB protein and peptides derived  
 CC from it are used in vaccine compositions for generating a protective  
 CC immune response against group A Streptococcus. Antibodies against GRAB  
 CC are useful for treating Streptococcus pyogenes infections. The protein  
 CC is also useful for purifying alpha2M from a sample. The present sequence  
 CC is a fragment of GRAB protein from S. pyogenes strain SF370 corresponding  
 CC to residues 34-56. This fragment is capable of binding alpha2M  
 CC and useful in vaccine composition.  
 CC  
 XX Sequence 23 AA;  
 SQ  
 Query Match 15.4%; Score 120; DB 21; Length 23;  
 Best Local Similarity 100.0%; Pred. No. 8.7e-08;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VDSPIEQPRIIPNGGTLNLLGN 23  
 Db 1 VDSPIEQPRIIPNGGTLNLLGN 23  
 RESULT 21  
 AAY71048  
 ID AAY71048 standard; peptide; 25 AA.  
 XX  
 XX AAY71048;  
 AC  
 XX 29-AUG-2000 (first entry)  
 DT  
 XX Streptococcus pyogenes GRAB peptide EKL 24.  
 DE  
 XX GRAB; protein G related alpha2M binding protein; vaccine;  
 KW alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;  
 KW immune response; Streptococcus pyogenes infection; peptide EKL 24.  
 XX  
 XX Streptococcus pyogenes.  
 OS  
 XX WO200026240-A2.  
 PN 11-MAY-2000.  
 PD  
 XX 02-NOV-1999; 99WO-GB03631.  
 PF  
 XX 02-NOV-1998; 98GB-0023975.  
 PR  
 XX (ACTI-) ACTINOVA LTD.  
 PA  
 XX Bjorck LH, Rasmussen M;  
 PI  
 XX WPI; 2000-365572/31.  
 DR  
 XX

PT New alpha2M binding protein for generating a protective immune response  
 PT to group A streptococcus and purifying the binding protein -  
 XX  
 XX Example 9; Page 31; 67pp; English.  
 PS  
 XX The patent discloses a new family of proteins termed GRAB (protein G  
 CC related alpha2M binding protein) from Streptococcus pyogenes which have  
 CC the ability to bind alpha2-macroglobulin (alpha2M) and show homology to  
 CC protein G of group G Streptococcus. GRAB protein and peptides derived  
 CC from it are used in vaccine compositions for generating a protective  
 CC immune response against group A Streptococcus. Antibodies against GRAB  
 CC are useful for treating Streptococcus pyogenes infections. The protein  
 CC is also useful for purifying alpha2M from a sample. The present sequence  
 CC is a GRAB peptide EKL 24 useful in vaccine composition. It was used  
 CC to produce immunoglobulin G (IgG) antibodies specific for native GRAB  
 CC protein in sheep. The peptide has a cysteine insert at the C-terminal  
 CC for attachment to a hetero-bifunctional linker.  
 XX  
 XX Sequence 25 AA;  
 SQ  
 Query Match 14.8%; Score 115; DB 21; Length 25;  
 Best Local Similarity 100.0%; Pred. No. 4e-07;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 26 EKLALRNEERAIDELKKQAIEDKE 49  
 Db 1 EKLALRNEERAIDELKKQAIEDKE 24  
 RESULT 22  
 AAP94785  
 ID AAP94785 standard; protein; 235 AA.  
 XX  
 XX AAP94785;  
 AC  
 XX 04-JUL-1990 (first entry)  
 DT  
 XX Protein G variant.  
 DE  
 XX Protein G; immunoglobulin; Fc receptor; ds.  
 KW  
 XX Streptococcus sp.  
 OS  
 XX WO8810306-A.  
 PN 29-DEC-1988.  
 XX  
 PD 20-JUN-1988; 88WO-US02084.  
 XX  
 PF 19-JUN-1987; 87US-0063959.  
 XX  
 PR (GENE-) GENEX CORP.  
 XX  
 PA Fahnstock SR;  
 XX  
 PI WPI; 1989-023848/03.  
 XX  
 DR N-PSDB; AAN91099.  
 DR  
 XX Cloned protein G variant genes -  
 XX expressing proteins having immunoglobulin-binding properties of  
 PT protein G and derived from Streptococcus sp.  
 PT  
 XX Claim 17; Page 88; 116pp; English.  
 PS  
 XX Gene for protein G variant of non-pathogenic streptococcus sp. allowing  
 CC isolation of the protein and variants, useful as bacterial Fc receptors  
 CC eg in purification and detection of Abs., screening of hybridoma clones  
 CC and treatment of disease.  
 CC  
 XX Sequence 235 AA;  
 SQ  
 Query Match 13.5%; Score 105; DB 10; Length 235;  
 Best Local Similarity 69.7%; Pred. No. 0.0001;

XX AC AAY57611;  
XX DT 10-MAR-2000 (first entry)  
XX DE Streptococcus strain G 148 protein.  
XX KW Streptococcus strain G 148; protein G'; protein G primer; liposome;  
KW liposomal delivery complex; connecting moiety; antibody; FC region;  
KW diagnosis; therapeutic agent; antibiotic; antipressant; antiviral;  
KW antitumorigenic; cytokine; hormone; imaging agent; neurotransmitter;  
KW stimulant; cytotoxic agent; malignant cell.  
XX OS Streptococcus sp.  
XX PN WO9959643-A2.  
XX PD 25-NOV-1999.  
XX PF 19-MAY-1999; 99WO-US11177.  
XX PR 20-MAY-1998; 98US-0086347.  
XX PA (SDGS-) SDG INC.  
XX PI Lau JR;  
XX DR WPI; 2000-062383/05.  
XX DR N-PSDB; AAZ47931.  
XX PT New liposome constructs comprising a liposome connected to an antibody,  
PT used, e.g. for delivery of cytotoxic agents to malignant cells -  
XX PS Disclosure: Fig 1; 32pp; English.  
XX CC The present invention describes a construct for connecting an antibody  
CC or antibody fragment to a liposome. The construct comprises protein G'  
CC and a linking moiety for connecting the protein G' to the liposome.  
CC Also described are: (1) a liposomal delivery vehicle comprising: (a) a  
CC liposome; and (b) a connecting moiety connected to the liposome, which  
CC specifically binds the FC region of an antibody, for connecting the  
CC antibody to the liposome; and (2) forming a liposomal carrier comprising:  
CC (a) providing liposomal components having binding moieties, for forming  
CC a liposome; (b) providing a construct comprising a linking moiety and a  
CC connecting moiety bound together; and (c) combining the liposomal  
CC components and the construct and sonicating the combination so that the  
CC binding moieties are exposed on a surface of the liposome, for binding to  
CC the linking moiety, and to facilitate the binding between them. The  
CC products can be used for the delivery of diagnostic or therapeutic  
CC agents. The liposomes may contain or may be associated with a diagnostic  
CC or therapeutic agent, e.g. antibiotics, antipressants,  
CC antitumorigenics, antivirals, cytokines, hormones, imaging agents,  
CC neurotransmitters, or stimulants. They can be used particularly for the  
CC delivery of cytotoxic agents to malignant cells. The protein G'  
CC connecting moiety provides a liposomal delivery complex having improved  
CC targeting efficiency. As a result of the binding between protein G' and  
CC the FC region of antibodies, protein G' shields the FC regions of the  
CC attached antibodies from non-specific binding to cell-surfaces, other  
CC proteins, and anatomical structures. The present sequence represents  
CC an unidentified protein encoded by the same sequence which encodes  
CC protein G'.  
XX SQ Sequence 103 AA;

Query Match 18.3%; Score 142; DB 21; Length 103;  
Best Local Similarity 57.9%; Pred. No. 1e-09;  
Matches 33; Conservative 3; Mismatches 21; Indels 0; Gaps 0;

QY 1 VDSPIEQPRIIPNGGTLTNLGNAPKELALRNEERAIDELKKQAIEDKATTAIEAA 57  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 34 VDSPIEDTPIIRNGGBELNLLGNSETTLALRNEESATADLTAAAVADTVAAAAAENA 90

RESULT 19  
AAY71038  
ID AAY71038 standard; peptide; 28 AA.  
XX AC AAY71038;  
XX DT 29-AUG-2000 (first entry)  
XX DE Streptococcus pyogenes strain SF370 GRAB protein fragment #3.  
XX KW GRAB protein; protein G related alpha2M binding protein; vaccine;  
KW alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;  
KW immune response; Streptococcus pyogenes infection.  
XX OS Streptococcus pyogenes.  
XX PN WO200026240-A2.  
XX PD 11-MAY-2000.  
XX PF 02-NOV-1999; 99WO-GB03631.  
XX PR 02-NOV-1998; 98GB-0023975.  
XX PA (ACTI-) ACTINOVA LTD.  
XX PI Bjorck LH, Rasmussen M;  
XX DR WPI; 2000-365572/31.  
XX PT New alpha2M binding protein for generating a protective immune response  
PT to group A streptococcus and purifying the binding protein -  
XX PS Claim 3; Page 56; 67pp; English.  
XX CC The patent discloses a new family of proteins termed GRAB (protein G  
CC related alpha2M binding protein) from Streptococcus pyogenes which have  
CC the ability to bind alpha2-macroglobulin (alpha2M) and show homology to  
CC protein G of group G Streptococcus. GRAB protein and peptides derived  
CC from it are used in vaccine compositions for generating a protective  
CC immune response against group A Streptococcus. Antibodies against GRAB  
CC are useful for treating Streptococcus pyogenes infections. The protein  
CC is also useful for purifying alpha2M from a sample. The present sequence  
CC is one of the repeat regions of GRAB protein from  
CC S. pyogenes strain SF370 corresponding to residues 92-119. This fragment  
CC is useful in vaccine composition.  
XX SQ Sequence 28 AA;

Query Match 16.6%; Score 129; DB 21; Length 28;  
Best Local Similarity 100.0%; Pred. No. 8.6e-09;  
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 59 SDALALADQTDALQSEEAAYVVKADNAA 86  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 1 SDALALADQTDALQSEEAAYVVKADNAA 28

RESULT 20  
AAY71036  
ID AAY71036 standard; peptide; 23 AA.  
XX AC AAY71036;  
XX DT 29-AUG-2000 (first entry)  
XX DE Streptococcus pyogenes strain SF370 GRAB protein fragment #1.  
XX KW GRAB protein; protein G related alpha2M binding protein; vaccine;  
KW alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;  
KW immune response; Streptococcus pyogenes infection.  
XX OS Streptococcus pyogenes.

FT Region /note= "involved in IgG binding activity"  
 FT 358..372 /label= b  
 FT /note= "linking region"  
 FT Active-site 373..427 /label= B3  
 FT /note= "involved in IgG binding activity"  
 FT Misc-difference 376 /note= "corresponds to CCT codon"  
 FT 428..442 /label= b  
 FT Region /note= "linking region"  
 FT Active-site 443..497 /label= B2  
 FT /note= "involved in IgG binding activity"  
 FT Misc-difference 466 /note= "corresponds to ACT codon"  
 FT 531..535 /label= C1  
 FT Region /label= C2  
 FT 536..540 /label= C2  
 FT Region /label= C3  
 FT 541..545 /label= C3  
 FT Region /label= C4  
 FT 546..550 /label= C4  
 FT Region /label= C5  
 FT 551..555 /label= C5  
 FT Misc-difference 592 /note= "corresponds to GAA codon"

XX US5312901-A.  
 XX 17-MAY-1994.  
 PD  
 XX 14-FEB-1986; 86US-0829354.  
 XX 14-FEB-1986; 86US-0829354.  
 PR 23-APR-1986; 86US-0854887.  
 PR 19-JUN-1987; 87US-0063959.  
 PR 20-JUN-1988; 88US-0209236.  
 PR 19-JUN-1990; 90US-0540169.  
 PR 21-APR-1992; 92US-0871539.  
 XX (PHAA ) PHARMACIA LKB BIOTECHNOLOGY AB.  
 XX Fahnestock SR;  
 PI WPI; 1994-159179/19.  
 DR N-PSDB; AAQ75036.  
 XX  
 PT New recombinant streptococcal protein G variants - useful for  
 PT antibody detection and purification and for therapy  
 PS Example 5; Fig 9; 48pp; English.

XX A 2.4kb HindIII fragment containing the entire coding sequence for  
 CC Protein G was isolated from Streptococcus GX7805 using the 1.9kb  
 CC Protein G coding sequence from Streptococcus GX7809. The Protein G  
 CC has IgG-binding activity which has been localised to the B repeating  
 CC structure. Streptococcal Protein G variants comprising the B domains  
 CC are claimed.

XX Sequence 593 AA;  
 Query Match 19.9%; Score 155; DB 15; Length 593;  
 Best Local Similarity 56.1%; Pred. No. 2.2e-10;  
 Matches 37; Conservative 4; Mismatches 25; Indels 0; Gaps 0;

QY 1 VDSPICQPIIPNGGTLTNLLGNAPKALRNEERAIDELKKQAIEDKENTTAIEAASSD 60  
 DB 34 VDSPIEDTPIIRNGGELTNLLGNSETTLALRNEESATADLTAAAVADTVAAAAAENAGAA 93  
 QY 61 ALEALA 66

Db 94 AWEAAA 99

## RESULT 17

AAR10005

ID AAR10005 standard; protein; 594 AA.

XX AAR10005;

XX 13-MAR-1991 (first entry)

XX Streptococcus GX7805 protein G.

XX Immunoglobulins; Ig.

XX Streptococcus sp GX7805.

XX Key Location/Qualifiers

FT Active-site 304..358

FT /label= Active Site B1

FT Active-site 374..428

FT /label= Active Site B3

FT Active-site 444..498

FT /label= Active Site B2

XX US4977247-A.

PN 11-DEC-1990.

XX 19-MAY-1989; 89US-0354264.

XX 19-MAY-1989; 89US-0354264.

PR 14-FEB-1986; 86US-0829354.

PR 23-APR-1986; 86US-0854887.

PR 17-FEB-1987; 87WO-US00329.

PR 19-JUN-1987; 87US-0063959.

PR 20-JUN-1988; 88US-0209236.

XX (GENE-) GENEX CORP.

XX Fahnestock SR, Lee T, Wroble MH;

XX WPI; 1991-006758/01.

DR Q-PSDB; Q10002.

XX

PT Immobilised protein G variants - used for detection, isolation

PT and purificn. immunoglobulin(s) and immunoglobulin fragments

XX Disclosure; Fig 9; 52pp; English.

XX Protein G gene product may be modified allowing the variant to

CC be immobilised and exhibit different binding profiles. The bound

CC protein is useful in purification and detection of Igs and fragments.

XX Sequence 594 AA;

Query Match 19.9%; Score 155; DB 12; Length 594;

Best Local Similarity 56.1%; Pred. No. 2.2e-10;

Matches 37; Conservative 4; Mismatches 25; Indels 0; Gaps 0;

QY 1 VDSPICQPIIPNGGTLTNLLGNAPKALRNEERAIDELKKQAIEDKENTTAIEAASSD 60

DB 34 VDSPIEDTPIIRNGGELTNLLGNSETTLALRNEESATADLTAAAVADTVAAAAAENAGAA 93

QY 61 ALEALA 66

Db 94 AWEAAA 99

XX

RESULT 18

AA57611

ID AA57611 standard; Protein; 103 AA.

```

XX PD 17-MAY-1994.
XX PF 14-FEB-1986; 86US-08293354.
XX PR 14-FEB-1986; 86US-08293354.
XX PR 23-APR-1986; 86US-0854887.
XX PR 19-JUN-1987; 87US-0063959.
XX PR 20-JUN-1988; 88US-0209236.
XX PR 19-JUN-1990; 90US-0540169.
XX PR 21-APR-1992; 92US-0871539.
XX PA (PHAA ) PHARMACIA LKB BIOTECHNOLOGY AB.
XX PI Fahnstock SR;
XX DR WPI; 1994-159179/19.
XX DR N-PSDB; AAQ64644.
XX PR New recombinant streptococcal protein G variants - useful for
XX antibody detection and purification and for therapy
XX PS Example 2; Fig 3 and Fig 8; 48pp; English.
XX CC A 1.9kb HindIII fragment containing the entire coding sequence for
XX Protein G was isolated from Streptococcus GX7809. The Protein G has
XX IgG-binding activity which has been localised to the B repeating
XX structure. Streptococcal Protein G variants comprising the B domains
XX are claimed.
XX SQ Sequence 448 AA;
Query Match 19.9%; Score 155; DB 15; Length 448;
Best Local Similarity 56.1%; Pred. No. 1.5e-10;
Matches 37; Conservative 4; Mismatches 25; Indels 0; Gaps 0;
QY 1 VDSPTEQPRIIPNGTTLNLGNAPKALRNEERAIDELKKQAIEDKEATTAIEAASSD 60
DB 34 VDSPIEDPIIRNGGELTNLGNSETTLALRNEESATADLTAAAVADTVAAAAAENAGAA 93
QY 61 ALEALA 66
DB 94 AWEAAA 99
RESULT 15
AAR07014
XX ID AAR07014 standard; protein; 593 AA.
XX AC AAR07014;
XX DT 17-JAN-1991 (first entry)
XX DE Protein G variant with three active sites.
XX KW Immunoglobulin.
XX OS Streptococcus sp. Lancefield Group G strain.
XX FH Key Location/Qualifiers
XX FT Active-site 303..372
XX FT Active-site /label=B1
XX FT Active-site 373..427
XX FT Active-site /label=B3
XX FT Active-site 443..497
XX FT Active-site /label=B2
XX PN US4956296-A.
XX PD 11-SEP-1990.
XX PF 20-JUN-1988; 88US-0209236.
XX

```

```

PR 20-JUN-1988; 88US-0209236.
PR 14-FEB-1986; 86US-08293354.
PR 23-APR-1986; 86US-0854887.
PR 17-FEB-1987; 87WO-US00329.
PR 19-JUN-1987; 87US-0063959.
XX (GENE-) GENEX CORP.
XX PI Fahnstock SR;
XX DR WPI; 1990-297491/39.
XX DR N-PSDB; AAQ06019.
XX PR Recombinant Protein G variants - obtd. using a cloned gene
XX encoding Protein G from Streptococcus sp., used for binding
XX immunoglobulin.
XX PS Disclosure; Fig 9; 48pp; English.
XX CC Fragments and variants of the sequence are claimed esp. where
XX incorporated into a non-pathogenic host eg. E.coli, and expressed
XX at high levels.
XX CC The variants have a higher binding efficiency and capacity for
XX immunoglobulin, and may be used for purifying, detecting and
XX isolating antibodies.
XX SQ Sequence 593 AA;
Query Match 19.9%; Score 155; DB 11; Length 593;
Best Local Similarity 56.1%; Pred. No. 2.2e-10;
Matches 37; Conservative 4; Mismatches 25; Indels 0; Gaps 0;
QY 1 VDSPTEQPRIIPNGTTLNLGNAPKALRNEERAIDELKKQAIEDKEATTAIEAASSD 60
DB 34 VDSPIEDPIIRNGGELTNLGNSETTLALRNEESATADLTAAAVADTVAAAAAENAGAA 93
QY 61 ALEALA 66
DB 94 AWEAAA 99
RESULT 16
AAR62944
XX ID AAR62944 standard; Protein; 593 AA.
XX AC AAR62944;
XX DT 10-JAN-1995 (first entry)
XX DE Streptococcus Protein G derived from strain GX7805.
XX KW Streptococcus Protein G; variant; IgG binding activity;
XX immunoglobulin; Lancefield Group G; bacterial Fc receptor.
XX OS Streptococcus sp. GX7805.
XX FH Key Location/Qualifiers
XX FT Domain 106..140
XX FT Domain /label=A1
XX FT Region 141..178
XX FT /label=a1
XX FT /note="linking region"
XX FT Domain 179..215
XX FT /label=A2
XX FT Region 216..253
XX FT /label=a2
XX FT /note="linking region"
XX FT Domain 254..290
XX FT /label=A3
XX FT Misc-difference 269
XX FT /note="corresponds to GGA codon"
XX FT Active-site 303..357
XX FT /label=B1

```

```

KW Immunoglobulin.
OS Streptococcus sp. Lancefield Group G strain.
XX
FH Key Location/Qualifiers
FT Active-site 228..282
FT Active-site /label=B1
FT Active-site 298..352
FT Active-site /label=B2
XX
PN US4956296-A.
XX
XX 11-SEP-1990.
XX
XX 20-JUN-1988; 88US-0209236.
XX
XX 20-JUN-1988; 88US-0209236.
PR 14-FEB-1986; 86US-0829354.
PR 23-APR-1986; 86US-0854887.
PR 17-FEB-1987; 87WO-US00329.
PR 19-JUN-1987; 87US-0063959.
XX
PA (GENE-) GENEX CORP.
XX
XX Fahnestock SR;
XX
DR WPI; 1990-297491/39.
DR N-PSDB; AAQ06018.
XX
PT Recombinant Protein G variants - obtd. using a cloned gene
PT encoding Protein G from Streptococcus sp., used for binding
PT Immunoglobulin.
XX
PS Disclosure; Fig 8a-c; 48pp; English.
XX
CC. Fragments and variants of the sequence are claimed esp. where
CC incorporated into a non-pathogenic host eg. E.coli, and expressed
CC at high levels.
CC The variants have a higher binding efficiency and capacity for
CC immunoglobulin, and may be used for purifying, detecting and
CC isolating antibodies.
XX.
SQ Sequence 448 AA;
Query Match 19.9%; Score 155; DB 11; Length 448;
Best Local Similarity 56.1%; Pred. No. 1.5e-10;
Matches 37; Conservative 4; Mismatches 25; Indels 0; Gaps 0;
QY 1 VDSPIEQPRIIPNGGTLTNLGNAPKALRNEERAIDELKKAIEDKKAATTAIEAASD 60
DB 34 VDSPIEDTPIIRNGGELTNLGNSETTLALRNEESATADLTAAAVADTVAAAAAENAGAA 93
QY 61 ALEALA 66
DB 94 AWEAAA 99
RESULT 13
AAR10004
ID AAR10004 standard; Protein; 448 AA.
XX
AC AAR10004;
XX
DT 13-MAR-1991 (first entry)
XX
DE Streptococcus GX7809 protein G.
XX
KW Immunoglobulins; Ig.
XX
OS Streptococcus sp GX7809.
XX
FH Key Location/Qualifiers
FT Active-site 228..282

```

```

FT Active-site /label= Active site B1
FT 298..352
FT Active-site /label= Active site B2
XX
PN US4977247-A.
XX
XX 11-DEC-1990.
XX
XX 19-MAY-1989; 89US-0354264.
XX
XX 19-MAY-1989; 89US-0354264.
PR 14-FEB-1986; 86US-0829354.
PR 23-APR-1986; 86US-0854887.
PR 17-FEB-1987; 87WO-US00329.
PR 19-JUN-1987; 87US-0063959.
PR 20-JUN-1988; 88US-0209236.
XX
XX (GENE-) GENEX CORP.
XX
XX Fahnestock SR, Lee T, Wroble MH;
XX
DR WPI; 1991-006758/01.
DR Q-PSDB; Q10001.
XX
PT Immobilised protein G variants - used for detection, isolation
PT and purificn. immunoglobulin(s) and immunoglobulin fragments
XX
PS Disclosure; Fig 8; 52pp; English.
XX
CC Protein G gene product may be modified allowing the variant to be
CC immobilised and exhibit different binding profiles. The bound
CC protein is useful in purification and detection of Igs and fragments.
XX
SQ Sequence 448 AA;
Query Match 19.9%; Score 155; DB 12; Length 448;
Best Local Similarity 56.1%; Pred. No. 1.5e-10;
Matches 37; Conservative 4; Mismatches 25; Indels 0; Gaps 0;
QY 1 VDSPIEQPRIIPNGGTLTNLGNAPKALRNEERAIDELKKAIEDKKAATTAIEAASD 60
DB 34 VDSPIEDTPIIRNGGELTNLGNSETTLALRNEESATADLTAAAVADTVAAAAAENAGAA 93
QY 61 ALEALA 66
DB 94 AWEAAA 99
RESULT 14
AAR53290
ID AAR53290 standard; Protein; 448 AA.
XX
AC AAR53290;
XX
XX 06-JAN-1995 (first entry)
XX
DE Streptococcus Protein G derived from strain GX7809.
XX
KW Streptococcus Protein G; variant; IgG binding activity;
KW Immunoglobulin; Lancefield Group G; bacterial Fc receptor.
XX
OS Streptococcus sp. GX7809.
XX
FH Key Location/Qualifiers
FT Active-site 228..282
FT Active-site /label= B1
FT Region 283..297
FT /label= b
FT /note= "linking region"
FT Active-site 298..352
FT Active-site /label= B2
XX
PN US5312901-A.

```

CC is a fragment of GRAB protein from S. pyogenes strain SF370 corresponding  
CC to residues 34-91. This fragment is capable of binding alpha2M  
CC and useful in vaccine composition.

XX Sequence 58 AA;

Query Match 36.7%; Score 285; DB 21; Length 58;  
Best Local Similarity 100.0%; Pred. No. 1.2e-27;  
Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VDSPIEQPRIIPNGGTLNLLGNAPEKALRNEERAIDELKQKAIEDKEATTAEAS 58  
Db 1 VDSPIEQPRIIPNGGTLNLLGNAPEKALRNEERAIDELKQKAIEDKEATTAEAS 58

## RESULT 10

AAP70493  
ID AAP70493 standard; protein; 448 AA.

XX AAP70493;

DT 06-MAR-1991 (first entry)

DE Protein G.

KW Protein G; antibody; Fc receptor;

OS Streptococcus Lancefield Group G strain.

FN Key Location/Qualifiers

FT Active-site 228..297

FT Active-site /label=active site B1

FT Active-site 298..352

FT Active-site /label=active site B2

PN W08705025-A.

XX 27-AUG-1987.

XX 17-FEB-1987; 87WO-US00329.

XX 22-APR-1986; 86US-0854997.

XX 14-FEB-1986; 86US-0829354.

XX (GENE-) GENEX CORP.

XX (FAHN/) FAHNSTOCK S R.

XX Fahnestock S;

XX WPI; 1987-250197/35.

XX N-PSDB; AAN70811.

PT Cloned Protein G gene - used for producing Protein G for  
PT detection and purificn. of antibodies and treatment of diseases

XX Disclosure; Fig. 3; 68pp; English.

XX Protein G expressed by inserting the gene into an expression  
CC vector. A second vector may also be used as a cryptic helper plasmid  
CC to stably maintain the first plasmid in the host cell. Bacterial Fc  
CC receptors such as protein G can be used to detect and purify  
CC antibodies, and in the treatment of disease. Fc receptors are useful  
CC to purify antibodies to be used in the purificn. of protein drugs and  
CC as therapeutics. High levels of protein G can be obt'd. in conditions  
CC favourable for isolation, using a non-pathogenic host. Suitable  
CC cloning vectors are lambda gt11, M13mp9 and pCX1066.

XX Sequence 448 AA;

Query Match 19.9%; Score 155; DB 8; Length 448;

Best Local Similarity 56.1%; Pred. No. 1.5e-10;

Matches 37; Conservative 4; Mismatches 25; Indels 0; Gaps 0;

QY 1 VDSPIEQPRIIPNGGTLNLLGNAPEKALRNEERAIDELKQKAIEDKEATTAEAS 60  
Db 34 VDSPIEQPRIIPNGGTLNLLGNAPEKALRNEESATADLTAAAVDTVAANAAGAA 93

QY 61 ALEALA 66

Db 94 AWEAAA 99

## RESULT 11

AAP95030

ID AAP95030 standard; protein; 448 AA.

XX AAP95030;

XX 04-JUL-1990 (first entry)

XX Protein G.

XX Protein G; immunoglobulin; Fc receptor; ds.

XX Streptococcus sp.

XX W08810306-A.

XX 29-DEC-1988.

XX 20-JUN-1988; 88WO-US02084.

XX 19-JUN-1987; 87US-0063959.

XX (GENE-) GENEX CORP.

XX Fahnestock SR;

XX WPI; 1989-023848/03.

XX N-PSDB; AAN91093.

XX Cloned protein G variant genes -  
PT expressing proteins having immunoglobulin-binding properties of  
PT protein G and derived from Streptococcus sp.

XX Disclosure; 116pp; English.

XX Protein G of non-pathogenic streptococcus and variants may be isolated,  
CC useful as bacterial Fc receptors eg in purification and detection of Abs.  
CC screening of hybridoma clones and treatment of disease.

XX Sequence 448 AA;

Query Match 19.9%; Score 155; DB 10; Length 448;

Best Local Similarity 56.1%; Pred. No. 1.5e-10;

Matches 37; Conservative 4; Mismatches 25; Indels 0; Gaps 0;

QY 1 VDSPIEQPRIIPNGGTLNLLGNAPEKALRNEERAIDELKQKAIEDKEATTAEAS 60

Db 34 VDSPIEQPRIIPNGGTLNLLGNAPEKALRNEESATADLTAAAVDTVAANAAGAA 93

QY 61 ALEALA 66

Db 94 AWEAAA 99

## RESULT 12

AAR07013

ID AAR07013 standard; protein; 448 AA.

XX AAR07013;

XX 17-JAN-1991 (first entry)

XX Protein G variant with two active sites.

PT New alpha2M binding protein for generating a protective immune response  
PT to group A streptococcus and purifying the binding protein -  
PS  
PS Claim 5; Page 61-62; 67pp; English.  
XX  
CC The patent discloses a new family of proteins termed GRAB (protein G  
CC related alpha2M binding protein) from Streptococcus pyogenes which have  
CC the ability to bind alpha2-macroglobulin (alpha2M) and show homology to  
CC protein G of group G Streptococcus. GRAB protein and peptides derived  
CC from it are used in vaccine compositions for generating a protective  
CC immune response against group A Streptococcus. Antibodies against GRAB  
CC are useful for treating Streptococcus pyogenes infections. The protein  
CC is also useful for purifying alpha2M from a sample. The present sequence  
CC is a partial GRAB protein from S. pyogenes strain AP49.  
CC The protein has alpha2M binding region and is useful in vaccine  
CC composition.  
XX  
SQ Sequence 271 AA;  
Query Match 69.0%; Score 536; DB 21; Length 271;  
Best Local Similarity 92.6%; Pred. NO. 9.4e-58;  
Matches 113; Conservative 2; Mismatches 7; Indels 0; Gaps 0;  
QY 1 VDSPIEQPRIIPNGGTLTNLLGNAPEKALRNEERAIDELKKQAIEDKEATTAEAAASD 60  
DB 4 VDSPIEQPRIIPNGGTLTNLLGNAPEKALRNEERAIDELKKQAIEDKEATTAEAAASD 63  
QY 61 ALEALADQTDALQSEEAHVVKADNAASDALEALADQTDALQSEEAHVVDNAASDAWEK 120  
DB 64 ALEALADQADALQSEEAHVVDNAASDALEALADQADALQSEEAHVVDNAAGDALEA 123  
QY 121 AA 122  
DB 124 LA 125  
RESULT 8  
AAV71043  
ID AAV71043 standard; Protein; 259 AA.  
XX  
AC AAV71043;  
XX  
DT 29-AUG-2000 (first entry)  
XX  
DE Streptococcus pyogenes strain KTL9 partial GRAB protein.  
XX  
KW GRAB protein; protein G related alpha2M binding protein; vaccine;  
KW alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;  
KW immune response; Streptococcus pyogenes infection.  
XX  
OS Streptococcus pyogenes.  
XX  
PN WO200026240-A2.  
XX  
PD 11-MAY-2000.  
XX  
PF 02-NOV-1999; 99WO-GB03631.  
XX  
PR 02-NOV-1998; 98GB-0023975.  
XX  
PA (ACTI-) ACTINOVA LTD.  
XX  
PI Bjorck LH, Rasmussen M;  
XX  
DR WPI; 2000-365572/31.  
DR N-PSDB; AAD00561.  
XX  
XX New alpha2M binding protein for generating a protective immune response  
PT to group A streptococcus and purifying the binding protein -  
PS Claim 5; Page 59-60; 67pp; English.  
XX  
CC The patent discloses a new family of proteins termed GRAB (protein G

CC related alpha2M binding protein) from Streptococcus pyogenes which have  
CC the ability to bind alpha2-macroglobulin (alpha2M) and show homology to  
CC protein G of group G Streptococcus. GRAB protein and peptides derived  
CC from it are used in vaccine compositions for generating a protective  
CC immune response against group A Streptococcus. Antibodies against GRAB  
CC are useful for treating Streptococcus pyogenes infections. The protein  
CC is also useful for purifying alpha2M from a sample. The present sequence  
CC is a partial GRAB protein from S. pyogenes strain KTL9.  
CC The protein has alpha2M binding region and is useful in vaccine  
CC composition.  
XX  
SQ Sequence 259 AA;  
Query Match 68.7%; Score 534; DB 21; Length 259;  
Best Local Similarity 91.8%; Pred. NO. 1.6e-57;  
Matches 112; Conservative 4; Mismatches 6; Indels 0; Gaps 0;  
QY 1 VDSPIEQPRIIPNGGTLTNLLGNAPEKALRNEERAIDELKKQAIEDKEATTAEAAASD 60  
DB 21 VDSPIEQPRIIPNGGTLTNLLGNAPEKALRNEERAIDELKKQAIEDKEATTAEAAASD 80  
QY 61 ALEALADQTDALQSEEAHVVKADNAASDALEALADQTDALQSEEAHVVDNAASDAWEK 120  
DB 81 ALEALADQADALQSEEAHVVDNAASDALEALADQADALQSEEAHVVDNAASDTLEA 140  
QY 121 AA 122  
DB 141 LA 142  
RESULT 9  
AAV71037  
ID AAV71037 standard; peptide; 58 AA.  
XX  
AC AAV71037;  
XX  
DT 29-AUG-2000 (first entry)  
XX  
DE Streptococcus pyogenes strain SF370 GRAB protein fragment #2.  
XX  
KW GRAB protein; protein G related alpha2M binding protein; vaccine;  
KW alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;  
KW immune response; Streptococcus pyogenes infection.  
XX  
OS Streptococcus pyogenes.  
XX  
PN WO200026240-A2.  
XX  
PD 11-MAY-2000.  
XX  
PF 02-NOV-1999; 99WO-GB03631.  
XX  
PR 02-NOV-1998; 98GB-0023975.  
XX  
PA (ACTI-) ACTINOVA LTD.  
XX  
PI Bjorck LH, Rasmussen M;  
XX  
DR WPI; 2000-365572/31.  
XX  
PT New alpha2M binding protein for generating a protective immune response  
PT to group A streptococcus and purifying the binding protein -  
PS Claim 2; Page 55; 67pp; English.  
XX  
CC The patent discloses a new family of proteins termed GRAB (protein G  
CC related alpha2M binding protein) from Streptococcus pyogenes which have  
CC the ability to bind alpha2-macroglobulin (alpha2M) and show homology to  
CC protein G of group G Streptococcus. GRAB protein and peptides derived  
CC from it are used in vaccine compositions for generating a protective  
CC immune response against group A Streptococcus. Antibodies against GRAB  
CC are useful for treating Streptococcus pyogenes infections. The protein  
CC is also useful for purifying alpha2M from a sample. The present sequence



PD 11-MAY-2000.  
XX  
PF 02-NOV-1999; 99WO-GB03631.  
XX  
PR 02-NOV-1998; 98GB-0023975.  
XX  
PA (ACTI-) ACTINOVA LTD.  
XX  
PI Bjorck LH, Rasmussen M;  
XX  
DR WPI; 2000-365572/31.  
XX  
DR N-PSDB; AAD00562.  
XX  
PT New alpha2M binding protein for generating a protective immune response  
PT to group A streptococcus and purifying the binding protein .  
XX  
PS Claim 5; Page 60-61; 67pp; English.  
XX  
CC The patent discloses a new family of proteins termed GRAB (protein G  
CC related alpha2M binding protein) from Streptococcus pyogenes which have  
CC the ability to bind alpha2-macroglobulin (alpha2M) and show homology to  
CC protein G of group G Streptococcus. GRAB protein and peptides derived  
CC from it are used in vaccine compositions for generating a protective  
CC immune response against group A Streptococcus. Antibodies against GRAB  
CC are useful for treating Streptococcus pyogenes infections. The protein  
CC is also useful for purifying alpha2M from a sample. The present sequence  
CC is a partial GRAB protein from S. pyogenes strain AP1.  
CC The protein has alpha2M binding region and is useful in vaccine  
CC composition.  
XX  
SQ Sequence 155 AA;  
Query Match 96.5%; Score 750; DB 21; Length 155;  
Best Local Similarity 99.4%; Pred. No. 2e-84;  
Matches 154; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
OY 2 DSPTEQRIIPNGGTLNLLGNAPEKALRNEERAIDELKKQAIEDKEATTAEASDDA 61  
DB 1 DSPTEQRIIPNGGTLNLLGNAPEKALRNEERAIDELKKQAIEDKEATTAEASDDA 60  
OY 62 LEALADQTDALQSEAAVVKADNAASDALEALADQTDALQSEAAEVVQSDNAASDAWEKA 121  
DB 61 LEALADQTDALQSEAAVVKADNAASDALEALADQTDALQSEAAEVVQSDNAASDAWEKA 120  
OY 122 ATPIALDVKKTKDTPVVKKEERQNVNLTPTTGE 156  
DB 121 ATPIALDVKKTKDTPVVKKEERQNVNLTPTTGE 155  
RESULT 6  
AAV71040  
ID AAV71040 standard; peptide; 141 AA.  
XX  
AC AAV71040;  
XX  
DT 29-AUG-2000 (first entry)  
XX  
DE Streptococcus pyogenes strain SF370 GRAB protein fragment #4.  
XX  
KW GRAB protein; protein G related alpha2M binding protein; vaccine;  
KW alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;  
KW immune response; Streptococcus pyogenes infection.  
XX  
OS Streptococcus pyogenes.  
XX  
PN WO200026240-A2.  
XX  
PD 11-MAY-2000.  
XX  
PF 02-NOV-1999; 99WO-GB03631.  
XX  
PR 02-NOV-1998; 98GB-0023975.  
XX

PA (ACTI-) ACTINOVA LTD.  
XX  
PI Bjorck LH, Rasmussen M;  
XX  
DR WPI; 2000-365572/31.  
XX  
PT New alpha2M binding protein for generating a protective immune response  
PT to group A streptococcus and purifying the binding protein .  
XX  
PS Claim 5; Page 57; 67pp; English.  
XX  
CC The patent discloses a new family of proteins termed GRAB (protein G  
CC related alpha2M binding protein) from Streptococcus pyogenes which have  
CC the ability to bind alpha2-macroglobulin (alpha2M) and show homology to  
CC protein G of group G Streptococcus. GRAB protein and peptides derived  
CC from it are used in vaccine compositions for generating a protective  
CC immune response against group A Streptococcus. Antibodies against GRAB  
CC are useful for treating Streptococcus pyogenes infections. The protein  
CC is also useful for purifying alpha2M from a sample. The present sequence  
CC is a fragment of GRAB protein from S. pyogenes strain SF370 corresponding  
CC to residues 34-174. This fragment is devoid of the trans-membrane  
CC and cell wall anchor regions. It is useful in vaccine composition.  
XX  
SQ Sequence 141 AA;  
Query Match 87.8%; Score 682; DB 21; Length 141;  
Best Local Similarity 100.0%; Pred. No. 4.4e-76;  
Matches 141; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 VDSPIEQRIIPNGGTLNLLGNAPEKALRNEERAIDELKKQAIEDKEATTAEASDD 60  
DB 1 VDSPIEQRIIPNGGTLNLLGNAPEKALRNEERAIDELKKQAIEDKEATTAEASDD 60  
OY 61 ALEALADQTDALQSEAAVVKADNAASDALEALADQTDALQSEAAEVVQSDNAASDAWEK 120  
DB 61 ALEALADQTDALQSEAAVVKADNAASDALEALADQTDALQSEAAEVVQSDNAASDAWEK 120  
OY 121 AATPIALDVKKTKDTPVVKK 141  
DB 121 AATPIALDVKKTKDTPVVKK 141  
RESULT 7  
AAV71045  
ID AAV71045 standard; Protein; 271 AA.  
XX  
AC AAV71045;  
XX  
DT 29-AUG-2000 (first entry)  
XX  
DE Streptococcus pyogenes strain AP49 partial GRAB protein.  
XX  
KW GRAB protein; protein G related alpha2M binding protein; vaccine;  
KW alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;  
KW immune response; Streptococcus pyogenes infection.  
XX  
OS Streptococcus pyogenes.  
XX  
PN WO200026240-A2.  
XX  
PD 11-MAY-2000.  
XX  
PF 02-NOV-1999; 99WO-GB03631.  
XX  
PR 02-NOV-1998; 98GB-0023975.  
XX  
PA (ACTI-) ACTINOVA LTD.  
XX  
PI Bjorck LH, Rasmussen M;  
XX  
DR WPI; 2000-365572/31.  
DR N-PSDB; AAD00563.  
XX

FT Region 183..188  
FT /note= "consensus sequence for gram-positive  
FT surface cell wall anchored proteins."  
FT 193..217  
FT /label= Membrane\_spanning\_region  
PN WO200026240-A2.  
XX  
XX 11-MAY-2000.  
XX  
PF 02-NOV-1999; 99WO-GB03631.  
XX  
PR 02-NOV-1998; 98GB-0023975.  
XX  
PA (ACTI-) ACTINOVA LTD.  
XX  
PI Bjorck LH, Rasmussen M;  
XX  
DR WPI; 2000-365572/31.  
DR N-PSDB; AAD00559, AAD00560.  
XX  
XX New alpha2M binding protein for generating a protective immune response  
PT to group A streptococcus and purifying the binding protein -  
XX  
XX Claim 5; Fig 2b; 67pp; English.  
XX  
XX The patent discloses a new family of proteins termed GRAB (protein G  
CC related alpha2M binding protein) from Streptococcus pyogenes which have  
CC the ability to bind alpha2-macroglobulin (alpha2M) and show homology to  
CC protein G of group G Streptococcus. GRAB protein and peptides derived  
CC from it are used in vaccine compositions for generating a protective  
CC immune response against group A Streptococcus. Antibodies against GRAB  
CC are useful for treating Streptococcus pyogenes infections. The protein  
CC is also useful for purifying alpha2M from a sample. The present sequence  
CC is a full-length GRAB protein from S. pyogenes strain SF370.  
XX  
SQ Sequence 217 AA;  
Query Match 100.0%; Score 777; DB 21; Length 217;  
Best Local Similarity 100.0%; Pred. No. 1.4e-87;  
Matches 199; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 VDSPIEQPRIIPNGGTLNLGNAPKALRNEERAIDELKKAIEDKEATTAEIASSD 60  
Db 34 VDSPIEQPRIIPNGGTLNLGNAPKALRNEERAIDELKKAIEDKEATTAEIASSD 93  
QY 61 ALEALADQTDALQSEEAAYVVKADNAASDALEALADQTDALQSEEAAYVQSDNAASDAWEK 120  
Db 94 ALEALADQTDALQSEEAAYVVKADNAASDALEALADQTDALQSEEAAYVQSDNAASDAWEK 153  
QY 121 AATPIALDVKKTKDTPVVKKEERQNVNLTPTTGEESNP 159  
Db 154 AATPIALDVKKTKDTPVVKKEERQNVNLTPTTGEESNP 192  
RESULT 4  
AAAY71046  
ID AAY71046 standard; Protein; 167 AA.  
XX  
XX AAY71046;  
AC  
XX 29-AUG-2000 (first entry)  
DT  
DE Streptococcus pyogenes strain KTL3 partial GRAB protein.  
XX  
XX GRAB protein; protein G related alpha2M binding protein; vaccine;  
KW alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;  
KW immune response; Streptococcus pyogenes infection.  
XX  
OS Streptococcus pyogenes.  
XX  
PN WO200026240-A2.  
XX

PD 11-MAY-2000.  
XX  
XX 02-NOV-1999; 99WO-GB03631.  
XX  
PR 02-NOV-1998; 98GB-0023975.  
XX  
PA (ACTI-) ACTINOVA LTD.  
XX  
XX Bjorck LH, Rasmussen M;  
XX  
DR WPI; 2000-365572/31.  
DR N-PSDB; AAD00564.  
XX  
XX New alpha2M binding protein for generating a protective immune response  
PT to group A streptococcus and purifying the binding protein -  
XX  
XX Claim 5; Page 62-63; 67pp; English.  
XX  
XX The patent discloses a new family of proteins termed GRAB (protein G  
CC related alpha2M binding protein) from Streptococcus pyogenes which have  
CC the ability to bind alpha2-macroglobulin (alpha2M) and show homology to  
CC protein G of group G Streptococcus. GRAB protein and peptides derived  
CC from it are used in vaccine compositions for generating a protective  
CC immune response against group A Streptococcus. Antibodies against GRAB  
CC are useful for treating Streptococcus pyogenes infections. The protein  
CC is also useful for purifying alpha2M from a sample. The present sequence  
CC is a partial GRAB protein from S. pyogenes strain KTL3.  
CC The protein has alpha2M binding region and is useful in vaccine  
CC composition.  
XX  
SQ Sequence 167 AA;  
Query Match 97.8%; Score 760; DB 21; Length 167;  
Best Local Similarity 100.0%; Pred. No. 1.3e-85;  
Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 VDSPIEQPRIIPNGGTLNLGNAPKALRNEERAIDELKKAIEDKEATTAEIASSD 60  
Db 12 VDSPIEQPRIIPNGGTLNLGNAPKALRNEERAIDELKKAIEDKEATTAEIASSD 71  
QY 61 ALEALADQTDALQSEEAAYVVKADNAASDALEALADQTDALQSEEAAYVQSDNAASDAWEK 120  
Db 72 ALEALADQTDALQSEEAAYVVKADNAASDALEALADQTDALQSEEAAYVQSDNAASDAWEK 131  
QY 121 AATPIALDVKKTKDTPVVKKEERQNVNLTPTTGE 156  
Db 132 AATPIALDVKKTKDTPVVKKEERQNVNLTPTTGE 167  
RESULT 5  
AAAY71044  
ID AAY71044 standard; Protein; 155 AA.  
XX  
XX AAY71044;  
AC  
XX 29-AUG-2000 (first entry)  
DT  
DE Streptococcus pyogenes strain AP1 partial GRAB protein.  
XX  
XX GRAB protein; protein G related alpha2M binding protein; vaccine;  
KW alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;  
KW immune response; Streptococcus pyogenes infection.  
XX  
OS Streptococcus pyogenes.  
XX  
XX Key Location/Qualifiers  
FH Misc-difference 17  
FT /note= "Thr at position 18 of GRAB protein fragment  
FT (AAY71036) from S. pyogenes strain SF370 is replaced  
FT with Ile"  
XX  
PN WO200026240-A2.  
XX

XX The patent discloses a new family of proteins termed GRAB (protein G  
CC related alpha2M binding protein) from Streptococcus pyogenes which have  
CC the ability to bind alpha2-macroglobulin (alpha2M) and show homology to  
CC protein G of group G Streptococcus. GRAB protein and peptides derived  
CC from it are used in vaccine compositions for generating a protective  
CC immune response against group A Streptococcus. Antibodies against GRAB  
CC are useful for treating Streptococcus pyogenes infections. The protein  
CC is also useful for purifying alpha2M from a sample. The present sequence  
CC is a fragment of GRAB protein from S. pyogenes strain SF370 corresponding  
CC to residues 34-192 and devoid of the membrane spanning region. This  
CC fragment is useful in vaccine composition.

XX Sequence 159 AA;  
SQ Query Match 100.0%; Score 777; DB 21; Length 159;  
Best Local Similarity 100.0%; Pred. No. 9.9e-88;  
Matches 159; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VDSPTEQPRIPNGGTLTLNLGNAPKALRNEERAIIDELKKQATIEDKATTAIEAASD 60

Db 1 VDSPTEQPRIPNGGTLTLNLGNAPKALRNEERAIIDELKKQATIEDKATTAIEAASD 60

OY 61 ALEALADQTDALQSEEAAYVVKADNAASDALEALADQTDALQSEEAAYVVDNAASDAWEK 120

Db 61 ALEALADQTDALQSEEAAYVVKADNAASDALEALADQTDALQSEEAAYVVDNAASDAWEK 120

OY 121 AATPTALDVKKTKDKTPVVKKEERQNVNTLPTTGESNP 159

Db 121 AATPTALDVKKTKDKTPVVKKEERQNVNTLPTTGESNP 159

## RESULT 2

AAV71039

ID AAV71039 standard; Protein: 184 AA.

XX AC AAV71039;

XX DT 29-AUG-2000 (first entry)

XX DE Streptococcus pyogenes strain SF370 mature GRAB protein.

XX GRAB protein; protein G related alpha2M binding protein; vaccine;  
KW alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;  
KW immune response; Streptococcus pyogenes infection.

XX OS Streptococcus pyogenes.

FH Key Location/Qualifiers

FT Binding-site 1..58

FT Region /label= alpha2-macroglobulin\_binding\_site

FT Region 59..86

FT Region /label= R1

FT Region /note= "repeat region"

FT Region 87..114

FT Region /label= R2

FT Region /note= "repeat region"

FT Region 115..159

FT Region /label= Cell\_wall\_spanning\_region

FT Region 150..155

FT Region /note= "consensus sequence for gram-positive

FT Region surface cell wall anchored proteins"

FT Region 160..184

FT Region /label= Membrane\_spanning\_region

XX WO200026240-A2.

XX PD 11-MAY-2000.

XX PF 02-NOV-1999; 99WO-GB03631.

XX PR 02-NOV-1998; 98GB-0023975.

XX

(ACTI-) ACTINOVA LTD.

XX Bjorck LH, Rasmussen M;

XX WPI: 2000-365572/31.

XX New alpha2M binding protein for generating a protective immune response

XX to group A streptococcus and purifying the binding protein -

XX Claim 5; Page 56; 67pp; English.

XX The patent discloses a new family of proteins termed GRAB (protein G  
CC related alpha2M binding protein) from Streptococcus pyogenes which have  
CC the ability to bind alpha2-macroglobulin (alpha2M) and show homology to  
CC protein G of group G Streptococcus. GRAB protein and peptides derived  
CC from it are used in vaccine compositions for generating a protective  
CC immune response against group A Streptococcus. Antibodies against GRAB  
CC are useful for treating Streptococcus pyogenes infections. The protein  
CC is also useful for purifying alpha2M from a sample. The present sequence  
CC is a mature GRAB protein from S. pyogenes strain SF370 without the  
CC signal sequence. This sequence is capable of binding alpha2M  
CC and useful in vaccine composition.

SQ Sequence 184 AA;

Query Match 100.0%; Score 777; DB 21; Length 184;

Best Local Similarity 100.0%; Pred. No. 1.2e-87;

Matches 159; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VDSPTEQPRIPNGGTLTLNLGNAPKALRNEERAIIDELKKQATIEDKATTAIEAASD 60

Db 1 VDSPTEQPRIPNGGTLTLNLGNAPKALRNEERAIIDELKKQATIEDKATTAIEAASD 60

OY 61 ALEALADQTDALQSEEAAYVVKADNAASDALEALADQTDALQSEEAAYVVDNAASDAWEK 120

Db 61 ALEALADQTDALQSEEAAYVVKADNAASDALEALADQTDALQSEEAAYVVDNAASDAWEK 120

OY 121 AATPTALDVKKTKDKTPVVKKEERQNVNTLPTTGESNP 159

Db 121 AATPTALDVKKTKDKTPVVKKEERQNVNTLPTTGESNP 159

## RESULT 3

AAV71042

ID AAV71042 standard; Protein: 217 AA.

XX AC AAV71042;

XX DT 29-AUG-2000 (first entry)

XX DE Streptococcus pyogenes strain SF370 full-length GRAB protein.

XX GRAB protein; protein G related alpha2M binding protein; vaccine;  
KW alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;  
KW immune response; Streptococcus pyogenes infection.

XX OS Streptococcus pyogenes.

FH Key Location/Qualifiers

FT Peptide 1..33

FT Protein /label= Signal\_sequence

FT Binding-site 34..217

FT Binding-site /label= Mature\_GRAB\_protein

FT Domain 34..91

FT Domain /label= alpha2M\_binding\_site

FT Region 34..68

FT Region /note= "shows homology to E domain of protein G"

FT Region 92..119

FT Region /label= Repeat\_region\_1

FT Region 120..147

FT Region /label= Repeat\_region\_2

FT Region 148..192

FT Region /label= Cell\_wall\_spanning\_region

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2002, 00:48:06 ; Search time 64.6203 Seconds  
(without alignments)  
273.300 Million cell updates/sec

Title: US-09-847-539A-6

Perfect score: 777

Sequence: 1 VDSPIEQPIIPNGTTLNL.....KKERQNVNTLPTTGESNP 159

Scoring table: BLOSUM62

Gapop 60.0 , Gapext 60.0

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_032802.\*  
1: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	777	100.0	159	21	AAV71041 Streptococcus pyog
2	777	100.0	184	21	AAV71039 Streptococcus pyog
3	777	100.0	217	21	AAV71042 Streptococcus pyog
4	760	97.8	167	21	AAV71046 Streptococcus pyog
5	750	96.5	155	21	AAV71044 Streptococcus pyog
6	682	87.8	141	21	AAV71040 Streptococcus pyog
7	536	69.0	271	21	AAV71045 Streptococcus pyog
8	534	68.7	259	21	AAV71043 Streptococcus pyog
9	285	36.7	58	21	AAV71037 Streptococcus pyog
10	155	19.9	448	8	AAV70493 Protein G. Strept
11	155	19.9	448	10	AAV95030 Protein G. Strept

12	155	19.9	448	11	AAV70113 Protein G variant
13	155	19.9	448	12	AAV10004 Streptococcus GX78
14	155	19.9	448	15	AAV53290 Streptococcus prot
15	155	19.9	593	11	AAV70104 Protein G variant
16	155	19.9	593	15	AAV62944 Streptococcus prot
17	155	19.9	594	12	AAV10005 Streptococcus GX78
18	142	18.3	103	21	AAV57611 Streptococcus stra
19	129	16.6	28	21	AAV71038 Streptococcus pyog
20	120	15.4	23	21	AAV71036 Streptococcus pyog
21	115	14.8	25	21	AAV71048 Streptococcus pyog
22	105	13.5	235	10	AAV94785 Protein G variant.
23	105	13.5	235	11	AAV70004 Protein G variant.
24	105	13.5	265	12	AAV10011 Type 4 GX7809 prot
25	105	13.5	265	15	AAV53294 IgG-binding Strept
26	97	12.5	20	21	AAV71051 Streptococcus pyog
27	96	12.4	19	21	AAV71047 Streptococcus pyog
28	93	12.0	422	22	ABG16038 Novel human diagno
29	93	12.0	459	22	ABG20309 Novel human diagno
30	91	11.7	413	16	AAV71928 S. dysgalactiae MA
31	91	11.7	413	16	AAV71670 S. dysgalactiae MA
32	90	11.6	480	8	AAV70468 Sequence of polype
33	86	11.1	19	21	AAV71049 Streptococcus pyog
34	86	11.1	1107	22	ABG18247 Novel human diagno
35	85	10.9	664	16	AAV71929 S. dysgalactiae MI
36	84	10.8	489	22	ABV58655 Drosophila melanog
37	83	10.7	46	16	AAV71125 SG-3. Synthetic.
38	83	10.7	168	21	AAV20163 Arabidopsis thalia
39	83	10.7	180	21	AAV20162 Arabidopsis thalia
40	83	10.7	208	16	AAV71128 Synthetic protein
41	83	10.7	208	16	AAV71127 Synthetic protein
42	82	10.6	18	21	AAV71050 Streptococcus pyog
43	81	10.4	1022	22	ABG25650 Novel human diagno
44	76	9.8	3248	17	AAV99795 Kinetochore protei
45	75	9.7	429	22	AAV01027 CFE 30 protein seq

#### ALIGNMENTS

RESULT 1

AAV71041

ID AAV71041 standard; peptide: 159 AA.

AC AAV71041;

DT 29-AUG-2000 (first entry)

XX Streptococcus pyogenes strain SF370 GRAB protein fragment #5.

DE GRAB protein; protein G related alpha2M binding protein; vaccine;

KW alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;

KW Immune response; Streptococcus pyogenes infection.

XX Streptococcus pyogenes.

XX WO200026240-42.

XX 11-MAY-2000.

XX 02-NOV-1999

XX 99WO-GB03631.

XX 02-NOV-1998; 98GB-0023975.

XX (ACTI-) ACTINOVA LTD.

XX Bjorck LH, Rasmussen M;

XX WPI; 2000-365572/31.

XX / New alpha2M binding protein for generating a protective immune response

PT to group A streptococcus and purifying the binding protein

XX Claim 5; Page 57-58; 67pp; English.

PS

```
OS Balanus amphitrite.
XX
FH Key Location/Qualifiers
FT CDS 1..1371
FT /*tag= a
FT /product= sixth_adhesion_protein
XX
XX JP11332573-A.
XX
XX 07-DEC-1999.
XX
XX 29-MAY-1998; 98JP-0149138.
XX
XX 29-MAY-1998; 98JP-0149138.
XX
XX (KAIY-) KAIYO BIOTECHNOLOGY KENKYUSHO KK.
XX
XX WPI; 2000-091356/08.
XX P-PSDB; AAY67238.
XX
XX Barnacle sixth adhesion protein gene - useful as a raw material for
XX adhesives
XX
XX Example 4; Page 4-5; 8pp; Japanese.
XX
XX This is the coding sequence of the barnacle sixth adhesion protein. The
XX protein was isolated from the cement secreted from the shell of
XX barnacles. The protein can be used as an adhesive in water, and in wet
XX environments.
XX
XX Sequence 1773 BP; 429 A; 499 C; 524 G; 321 T; 0 other;
XX
XX
XX Alignment Scores:
XX Pred. No.: 333 Length: 1773
XX Score: 45.00 Matches: 11
XX Percent Similarity: 53.85% Conservative: 3
XX Best Local Similarity: 42.31% Mismatches: 12
XX Query Match: 34.88% Indels: 0
XX DB: 21 Gaps: 0
XX
XX US-09-847-539A-6_COPY_59_86 (1-28) x AA256162 (1-1773)
XX
XX Qy 3 AlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVal 22
XX ||||||| : : ||| : : ||||||| : :
XX Db 599 GCATTGGAGACTGTGCTGGATCCTCCGGGAACGTCACGTCCTCGAGCGCGGTGTTG 540
XX
XX Qy 23 LysAlaAspAsnAlaAla 28
XX |||
XX Db 539 AGCGCCTCAGAAAGCTGCG 522
XX
XX Search completed: October 13, 2002, 02:41:30
XX Job time : 1028.43 secs
```



```

PD 02-DEC-1999.
XX
XX 21-MAY-1999; 99WO-AU00385.
XX
XX 21-MAY-1998; 98AU-0003634.
XX
XX (UNSY ) UNIV SYDNEY.
XX
XX Reeves PR, Wang L;
XX
XX WPI; 2000-072598/06.
XX
XX Novel nucleic acid molecule useful for the detection of flagellated
XX bacterial strains in food, faeces, etc.
XX
XX Claim 3; Page 200-201; 245pp; English.
XX
XX AA256331 to AA256398 represent nucleic acid molecules (I) encoding all
XX or part of an Escherichia coli flagellin protein except a protein
XX expressed by E. coli H1, H7, H12 or H48 type strains. The present
XX invention also describes a method of detecting the presence of E. coli
XX of a particular H serotype in a sample, comprising specifically
XX hybridising a nucleic acid, preferably at least a pair, derived from a
XX flagellating gene, specific for a particular flagellin gene associated
XX with the H serotype, to any E. coli in the sample which contain the gene,
XX and detecting any hybridised molecules, identifying the presence of that
XX serotype in the sample. (I) are useful for: (1) detecting the presence
XX of E. coli of H serotype in a sample by hybridising at least one or a
XX pair of (I) to any E. coli in the sample and detecting the hybridised
XX nucleic acid molecules; and (2) for detecting the presence of both O
XX and H-serotypes of E. coli by hybridising at least one or a pair of (I)
XX to any E. coli present in the sample and detecting the hybridised
XX nucleic acid molecules. (I) is particularly useful for detecting the
XX combination of O and H antigen. Hybridised (I) when using at least one
XX (I) is detected by southern blot analysis and, when using a pair of (I),
XX is detected by polymerase chain reaction (PCR). AA256399 to AA256420
XX represent primers used in the exemplification of the present invention.
XX
XX Sequence 1380 BP; 424 A; 299 C; 309 G; 348 T; 0 other;
XX
Alignment Scores:
Pred. No.: 243 Length: 1380
Score: 45.00 Matches: 9
Percent Similarity: 64.00% Conservative: 7
Best Local Similarity: 36.00% Mismatches: 9
Query Match: 34.88% Indels: 0
DB: 21 Gaps: 0
US-09-847-539A-6_COPY_59_86 (1-28) x AA256351 (1-1380)
QY 2 AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaVal 21
DB 1123 GATGCAGCAAAATCGTTGCAATCTACCACCAACCCGCTCGAAACTATCGACAAGCATTTG 1182
QY 22 ValLysAlaAspAsn 26
DB 1183 GCTAAAGTTGACAAT 1197
RESULT 37
AAZ56376
ID AA256376 standard; DNA; 1380 BP.
XX
XX AA256376;
XX
XX 17-MAR-2000 (first entry)
XX
XX Escherichia coli flagellin protein nucleotide sequence SEQ ID NO:46.
XX
XX Flagellin; fliC; antigen; detection; ds.
XX
XX Escherichia coli.
XX
XX WO9961458-A1.
XX

```

```

XX
XX 02-DEC-1999.
XX
XX 21-MAY-1999; 99WO-AU00385.
XX
XX 21-MAY-1998; 98AU-0003634.
XX
XX (UNSY ) UNIV SYDNEY.
XX
XX Reeves PR, Wang L;
XX
XX WPI; 2000-072598/06.
XX
XX Novel nucleic acid molecule useful for the detection of flagellated
XX bacterial strains in food, faeces, etc.
XX
XX Claim 3; Page 222; 245pp; English.
XX
XX AA256331 to AA256398 represent nucleic acid molecules (I) encoding all
XX or part of an Escherichia coli flagellin protein except a protein
XX expressed by E. coli H1, H7, H12 or H48 type strains. The present
XX invention also describes a method of detecting the presence of E. coli
XX of a particular H serotype in a sample, comprising specifically
XX hybridising a nucleic acid, preferably at least a pair, derived from a
XX flagellating gene, specific for a particular flagellin gene associated
XX with the H serotype, to any E. coli in the sample which contain the gene,
XX and detecting any hybridised molecules, identifying the presence of that
XX serotype in the sample. (I) are useful for: (1) detecting the presence
XX of E. coli of H serotype in a sample by hybridising at least one or a
XX pair of (I) to any E. coli in the sample and detecting the hybridised
XX nucleic acid molecules; and (2) for detecting the presence of both O
XX and H-serotypes of E. coli by hybridising at least one or a pair of (I)
XX to any E. coli present in the sample and detecting the hybridised
XX nucleic acid molecules. (I) is particularly useful for detecting the
XX combination of O and H antigen. Hybridised (I) when using at least one
XX (I) is detected by southern blot analysis and, when using a pair of (I),
XX is detected by polymerase chain reaction (PCR). AA256399 to AA256420
XX represent primers used in the exemplification of the present invention.
XX
XX Sequence 1380 BP; 419 A; 299 C; 313 G; 349 T; 0 other;
XX
Alignment Scores:
Pred. No.: 243 Length: 1380
Score: 45.00 Matches: 9
Percent Similarity: 64.00% Conservative: 7
Best Local Similarity: 36.00% Mismatches: 9
Query Match: 34.88% Indels: 0
DB: 21 Gaps: 0
US-09-847-539A-6_COPY_59_86 (1-28) x AA256376 (1-1380)
QY 2 AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaVal 21
DB 1123 GATGCAGCAAAATCGTTGCAATCTACCACCAACCCGCTCGAAACTATCGACAAGCATTTG 1182
QY 22 ValLysAlaAspAsn 26
DB 1183 GCTAAAGTTGACAAT 1197
RESULT 38
AAZ56340
ID AA256340 standard; DNA; 1383 BP.
XX
XX AA256340;
XX
XX 17-MAR-2000 (first entry)
XX
XX Escherichia coli flagellin protein nucleotide sequence SEQ ID NO:10.
XX
XX Flagellin; fliC; antigen; detection; ds.
XX
XX Escherichia coli.
XX
XX

```

PF 21-MAY-1999; 99WO-AU00385.  
 XX  
 PR 21-MAY-1998; 98AU-0003634.  
 XX  
 PA (UNSY ) UNIV SYDNEY.  
 XX  
 PI Reeves PR, Wang L;  
 XX  
 DR WPI; 2000-072598/06.  
 XX  
 XX Novel nucleic acid molecule useful for the detection of flagellated  
 PT bacterial strains in food, faeces, etc.  
 XX  
 PS Claim 3; Page 204; 245pp; English.  
 XX  
 CC AAZ56331 to AAZ56398 represent nucleic acid molecules (I) encoding all  
 CC or part of an Escherichia coli flagellin protein except a protein  
 CC expressed by E. coli H1, H7, H12 or H48 type strains. The present  
 CC invention also describes a method of detecting the presence of E. coli  
 CC of a particular H serotype in a sample, comprising specifically  
 CC hybridising a nucleic acid, preferably at least a pair, derived from a  
 CC flagellating gene, specific for a particular flagellin gene associated  
 CC with the H serotype, to any E. coli in the sample which contain the gene,  
 CC and detecting any hybridised molecules, identifying the presence of that  
 CC serotype in the sample. (I) are useful for: (1) detecting the presence  
 CC of E. coli of H serotype in a sample by hybridising at least one or a  
 CC pair of (I) to any E. coli in the sample and detecting the hybridised  
 CC nucleic acid molecules; and (2) for detecting the presence of both O  
 CC and H-serotypes of E. coli by hybridising at least one or a pair of (I)  
 CC to any E. coli present in the sample and detecting the hybridised  
 CC nucleic acid molecules. (I) is particularly useful for detecting the  
 CC combination of O and H antigen. Hybridised (I) when using at least one  
 CC (I) is detected by southern blot analysis and, when using a pair of (I),  
 CC is detected by polymerase chain reaction (PCR). AAZ56399 to AAZ56420  
 CC represent primers used in the exemplification of the present invention.  
 XX  
 SQ Sequence 1365 BP; 412 A; 289 C; 312 G; 352 T; 0 other;

Alignment Scores:  
 Pred. No.: 239 Length: 1365  
 Score: 45.00 Matches: 9  
 Percent Similarity: 64.00% Conservative: 7  
 Best Local Similarity: 36.00% Mismatches: 9  
 Query Match: 34.88% Indels: 0  
 DB: 21 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x AAZ56356 (1-1365)

QY 2 AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaVal 21  
 ||||| :|||:|  
 Db 1108 GATGCAGCAAAATCGTTGCAATCTACTACCAACCCGCTCGAAGCATTG 1167  
 ||||| :|||:|

QY 22 ValLysAlaAspAsn 26  
 |||||  
 Db 1168 GCTAAAGTTGACAAT 1182  
 |||||

RESULT 35  
 AAZ56343  
 ID AAZ56343 standard; DNA; 1368 BP.  
 XX  
 AC AAZ56343;  
 XX  
 DT 17-MAR-2000 (first entry)  
 XX  
 XX Escherichia coli flagellin protein nucleotide sequence SEQ ID NO:13.  
 DE  
 XX Flagellin; fliC; antigen; detection; ds.  
 KW  
 XX Escherichia coli.  
 OS  
 XX WO9961458-A1.  
 PN  
 XX 02-DEC-1999.  
 PD

XX 21-MAY-1999; 99WO-AU00385.  
 PF  
 XX  
 PR 21-MAY-1998; 98AU-0003634.  
 XX  
 PA (UNSY ) UNIV SYDNEY.  
 XX  
 PI Reeves PR, Wang L;  
 XX  
 DR WPI; 2000-072598/06.  
 XX  
 XX Novel nucleic acid molecule useful for the detection of flagellated  
 PT bacterial strains in food, faeces, etc.  
 XX  
 PS Claim 3; Page 195; 245pp; English.  
 XX  
 CC AAZ56331 to AAZ56398 represent nucleic acid molecules (I) encoding all  
 CC or part of an Escherichia coli flagellin protein except a protein  
 CC expressed by E. coli H1, H7, H12 or H48 type strains. The present  
 CC invention also describes a method of detecting the presence of E. coli  
 CC of a particular H serotype in a sample, comprising specifically  
 CC hybridising a nucleic acid, preferably at least a pair, derived from a  
 CC flagellating gene, specific for a particular flagellin gene associated  
 CC with the H serotype, to any E. coli in the sample which contain the gene,  
 CC and detecting any hybridised molecules, identifying the presence of that  
 CC serotype in the sample. (I) are useful for: (1) detecting the presence  
 CC of E. coli of H serotype in a sample by hybridising at least one or a  
 CC pair of (I) to any E. coli in the sample and detecting the hybridised  
 CC nucleic acid molecules; and (2) for detecting the presence of both O  
 CC and H-serotypes of E. coli by hybridising at least one or a pair of (I)  
 CC to any E. coli present in the sample and detecting the hybridised  
 CC nucleic acid molecules. (I) is particularly useful for detecting the  
 CC combination of O and H antigen. Hybridised (I) when using at least one  
 CC (I) is detected by southern blot analysis and, when using a pair of (I),  
 CC is detected by polymerase chain reaction (PCR). AAZ56399 to AAZ56420  
 CC represent primers used in the exemplification of the present invention.  
 XX  
 SQ Sequence 1368 BP; 419 A; 292 C; 312 G; 345 T; 0 other;

Alignment Scores:  
 Pred. No.: 240 Length: 1368  
 Score: 45.00 Matches: 9  
 Percent Similarity: 64.00% Conservative: 7  
 Best Local Similarity: 36.00% Mismatches: 9  
 Query Match: 34.88% Indels: 0  
 DB: 21 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x AAZ56343 (1-1368)

QY 2 AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaVal 21  
 ||||| :|||:|  
 Db 1111 GATGCAGCAAAATCGTTGCAATCTACTACCAACCCGCTCGAAGCATTG 1170  
 ||||| :|||:|

QY 22 ValLysAlaAspAsn 26  
 |||||  
 Db 1171 GCTAAAGTTGACAAT 1185  
 |||||

RESULT 36  
 AAZ56351  
 ID AAZ56351 standard; DNA; 1380 BP.  
 XX  
 AC AAZ56351;  
 XX  
 DT 17-MAR-2000 (first entry)  
 XX  
 XX Escherichia coli flagellin protein nucleotide sequence SEQ ID NO:21.  
 DE  
 XX Flagellin; fliC; antigen; detection; ds.  
 KW  
 XX Escherichia coli.  
 OS  
 XX WO9961458-A1.  
 PN  
 XX 02-DEC-1999.  
 PD



XX Claim 3; Page 190; 245pp; English.  
XX  
CC AA256331 to AA256398 represent nucleic acid molecules (1) encoding all  
CC or part of an Escherichia coli flagellin protein except a protein  
CC expressed by E. coli H1, H7, H12 or H48 type strains. The present  
CC invention also describes a method of detecting the presence of E. coli  
CC of a particular H serotype in a sample, comprising specifically  
CC hybridising a nucleic acid, preferably at least a pair, derived from a  
CC flagellating gene, specific for a particular flagellin gene associated  
CC with the H serotype, to any E. coli in the sample which contain the gene,  
CC and detecting any hybridised molecules, identifying the presence of that  
CC serotype in the sample. (1) are useful for: (1) detecting the presence  
CC of E. coli of H serotype in a sample by hybridising at least one or a  
CC pair of (1) to any E. coli in the sample and detecting the hybridised  
CC nucleic acid molecules; and (2) for detecting the presence of both O  
CC and H-serotypes of E. coli by hybridising at least one or a pair of (1)  
CC to any E. coli present in the sample and detecting the hybridised  
CC nucleic acid molecules. (1) is particularly useful for detecting the  
CC combination of O and H antigen. Hybridised (1) when using at least one  
CC (1) is detected by southern blot analysis and, when using a pair of (1),  
CC is detected by polymerase chain reaction (PCR). AA256399 to AA256420  
CC represent primers used in the exemplification of the present invention.  
XX  
SQ Sequence 417 BP; 120 A; 108 C; 95 G; 94 T; 0 other;

Alignment Scores:  
Pred. No.: 53.7 Length: 417  
Score: 45.00 Matches: 9  
Percent Similarity: 64.00% Conservative: 7  
Best Local Similarity: 36.00% Mismatches: 9  
Query Match: 34.88% Indels: 0  
DB: 21 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x AA256335 (1-417)

Qy 2 AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaVal 21  
Db 135 GATGAGCAAAATCGTTGCAATCTACCAACCGCTCGAAACTATCGACAAGCATTTG 194  
Qy 22 VallysAlaAspAsn 26  
Db 195 GCTAAAGTTGACAAAT 209

RESULT 33

AAF08476/c

ID AAF08476 standard; cDNA; 461 BP.

AC AAF08476;

DT 13-MAR-2001 (first entry)

DE Fusarium venenatum EST SEQ ID NO:999.

XX Multiple gene expression; filamentous fungal cell; EST;  
KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;  
KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;  
KW culture condition; environmental stress; spore morphogenesis;  
KW metabolic pathway engineering; catabolic pathway engineering; ss.

OS Fusarium venenatum.

XX WO200056762-A2.

PN 28-SEP-2000.

XX 22-MAR-2000; 2000WO-US07781.

XX 22-MAR-1999; 99US-0273623.

XX (NOVO ) NOVO NORDISK BIOTECH INC.

XX (NOVO ) NOVO NORDISK AS.

PI Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;  
XX WPI; 2000-594572/56.  
DR  
XX Monitoring differential expression of genes in filamentous fungal cells  
PT uses fluorescence-labeled nucleic acids isolated from the cells and a  
PT substrate of expressed sequence tags -  
XX  
PS Claim 86; Page 763-764; 3161pp; English.

CC The present invention describes a method for monitoring differential  
CC expression of genes in a first filamentous fungal (FF) cell relative to  
CC expression of the same genes in one or more second filamentous fungal  
CC cells. The method uses fluorescence-labeled nucleic acids isolated from  
CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs  
CC are used in the methods for monitoring differential expression of genes  
CC in a first filamentous fungal (FF) cell relative to expression of the  
CC same genes in one or more second filamentous fungal cells. Monitoring  
CC the global expression of genes from FF cells allows the production  
CC potential of the microorganisms to be improved. New genes may be  
CC discovered, possible functions of unknown open reading frames can be  
CC identified and gene copy number variation and stability can be  
CC monitored. The expression of genes can be used to study how FF cells  
CC adapt to changes in culture conditions, environmental stress, spore  
CC morphogenesis, recombination, metabolic or catabolic pathway  
CC engineering. Using ESTs provides several advantages over genomic or  
CC random cDNA clones including elimination of redundancy as one spot on an  
CC array equals one gene or open reading frame, and organisation of the  
CC microarrays based on function of the gene products to facilitate  
CC analysis of the results. AAF07478 to AAF11247 represents ESTs from  
CC Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus  
CC niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and  
CC AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are  
CC all specifically claimed in the present invention.

XX Sequence 461 BP; 117 A; 131 C; 122 G; 89 T; 2 other;

Alignment Scores:  
Pred. No.: 61 Length: 461  
Score: 45.00 Matches: 10  
Percent Similarity: 62.50% Conservative: 5  
Best Local Similarity: 41.67% Mismatches: 9  
Query Match: 34.88% Indels: 0  
DB: 21 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x AAF08476 (1-461)

Qy 5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVallysAla 24

Db 176 GAAGCACTCACGGCACAGGTTGAGTCGTACTTCGATAAGACCGCCTTGCTTGA 117

Qy 25 AspAsnAlaAla 28

Db 116 GACACGGCAGCT 105

RESULT 34

AAZ56356

ID AAZ56356 standard; DNA; 1365 BP.

XX AAZ56356;

AC AAZ56356;

DT 17-MAR-2000 (first entry)

DE Escherichia coli flagellin protein nucleotide sequence SEQ ID NO:26.

XX Flagellin; fliC; antigen; detection; ds.

OS Escherichia coli.

XX WO9961458-A1.

PN 02-DEC-1999.

XX



```

XX SQ Sequence 69936 BP; 16512 A; 19259 C; 17399 G; 16763 T; 3 other;
Alignment Scores:
Pred. No.: 2.28e+04 Length: 69936
Score: 46.00 Matches: 11
Percent Similarity: 65.38% Conservatives: 6
Best Local Similarity: 42.31% Mismatches: 9
Query Match: 35.66% Indels: 0
DB: 21 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x AAA81479 (1-69936)
Qy 3 AlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVal 22
Db 69296 GCTTTGGAGCGCGTTTGGAAAAAATCGAAGCGCGCAAAATCCGACGGCAGCATTTG 69237

Qy 23 LysAlaAspAsnAlaAla 28
Db 69236 GAAGCGGCTGAAGCGGCT 69219

RESULT 29
AAF21607
ID AAF21607 standard; DNA: 349980 BP.
AC AAF21607;
XX
XX DT 13-MAR-2001 (first entry)
XX DE Neisseria meningitidis B nucleotide sequence SEQ ID NO:108.
XX KW Neisseria meningitidis; Neisseria gonorrhoeae; immunogenic; vaccine;
XX KW diagnosis; antigen; detection; infection; gene therapy; antibacterial;
XX KW ds.
XX OS Neisseria meningitidis.
XX PN WO200066791-A1.
XX PD 09-NOV-2000.
XX PF 08-MAR-2000; 2000WO-US05928.
XX PR 30-APR-1999; 99US-0132068.
XX PR 08-OCT-1999; 99WO-US23573.
XX PR 28-FEB-2000; 2000GB-0004695.
XX PA (CHIR ) CHIRON CORP.
XX PA (GENO-) INST GENOMIC RES.
XX PI Pizza M, Hickey E, Peterson J, Tettelin H, Venter JC, Massignani V;
XX PI Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V, Rappuoli R;
XX PI Frazer CM, Grandi G;
XX WPI; 2000-647603/62.
XX
XX PT Neisseria meningitidis B full length genome sequence and open reading
XX PT frames are used to detect, treat and prevent Neisserial infections .
XX PS Claim 7; Appendix A; 692pp; English.
XX
CC The present invention describes the full length genome of
CC Neisseria meningitidis B (NMB). The sequences in AAF21544 and AAF21607
CC to AAF21613 represent fragments of the NMB genomic sequence, as the
CC sequence was too long to go in a record on its own it was split into 8
CC sequences which overlap each other at the beginning and end of each
CC sequence by 49980 bp (i.e. the last 49980 bp of AAF21544 is repeated at
CC the beginning of AAF21607, the last 49980 bp of AAF21607 are repeated at
CC the beginning of AAF21608, and so on). AAF21545 to AAF21588 encode the
CC Neisseria proteins given in AAB58550 to AAB58593, and AAF21589 to
CC AAF21606 represent PCR primers which are used in the exemplification of
CC the present invention. The NMB genome and fragments from it have
CC antibacterial activity, and can be used in vaccines and gene therapy.

```

---

```

CC Neisseria nucleic acids, proteins and/or antibodies which binds to the
CC proteins can be used in compositions for treating or preventing infection
CC due to Neisserial bacteria or as a diagnostic reagent for detecting the
CC presence of Neisserial bacteria or of antibodies raised to Neisserial
CC bacteria. Computers, computer memory, computer storage medium or computer
CC databases can be used in a search to identify open reading frames (ORFs)
CC or coding sequences within the NMB genome. The DNA sequences provide
CC further opportunities to find antigenic or immunogenic proteins which are
CC more effective in vaccines than the outer membrane proteins currently
CC used.
XX
XX SQ Sequence 349980 BP; 84410 A; 84863 C; 94187 G; 86520 T; 0 other;
Alignment Scores:
Pred. No.: 1.69e+05 Length: 349980
Score: 46.00 Matches: 11
Percent Similarity: 65.38% Conservatives: 6
Best Local Similarity: 42.31% Mismatches: 9
Query Match: 35.66% Indels: 0
DB: 21 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x AAF21607 (1-349980)
Qy 3 AlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVal 22
Db 111891 GCTTTGGAGCGCGTTTGGAAAAAATCGAAGCGCGCAAAATCCGACGGCAGCATTTG 111950

Qy 23 LysAlaAspAsnAlaAla 28
Db 111951 GAAGCGGCTGAAGCGGCT 111968

RESULT 30
AAH68525
ID AAH68525 standard; DNA: 349980 BP.
XX
XX AC AAH68525;
XX
XX DT 26-SEP-2001 (first entry)
XX DE C glutamicum coding sequence fragment SEQ ID NO: 7060.
XX KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
XX KW organic acid synthesis; ds.
XX OS Corynebacterium glutamicum.
XX PN EP1108790-A2.
XX PD 20-JUN-2001.
XX PF 18-DEC-2000; 2000EP-0127688.
XX PR 16-DEC-1999; 99JP-0377484.
XX PR 07-APR-2000; 2000JP-0159182.
XX PR 03-AUG-2000; 2000JP-0280988.
XX PA (KYOW ) KYOWA HAKKO KOGYO KK.
XX PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
XX PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX WPI; 2001-376931/40.
XX
XX PT Novel polynucleotides derived from Coryneform bacteria, for identifying
XX PT mutation point of a gene, measuring expression of a gene, analysing
XX PT expression profile or pattern of a gene and identifying homologous gene
XX PS Disclosure; SEQ ID NO: 7060; 246pp + Sequence Listing; English.
XX CC The present invention provides a number of nucleotide and protein
XX CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
XX CC are useful for identifying the mutation point of a gene derived from a

```

QY 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLys 23  
:::||||| ||| :::: ||| |||::: |||::: |||::: |||::: |||  
Db 1540 ATTGAGCCTTACCGACTGAGGAGGAGGCTTGAAGAGGGCGAGTCCGCAATCTACAAG 1599  
QY 24 AlaAspAsnAla 27  
:::|||||  
Db 1600 TTCGAGAACGCT 1611  
RESULT 27  
AAH65521/c  
ID AAH65521 standard; DNA; 2160 BP.  
XX  
AC AAH65521;  
DT 26-SEP-2001 (first entry)  
XX  
DE C glutamicum coding sequence fragment SEQ ID NO: 556.  
XX  
KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;  
KW organic acid synthesis; ds.  
XX  
OS Corynebacterium glutamicum.  
XX  
PN EP1108790-A2.  
XX  
PD 20-JUN-2001.  
XX  
PF 18-DEC-2000; 2000EP-0127688.  
XX  
PR 16-DEC-1999; 99JP-0377484.  
PR 07-APR-2000; 2000JP-0159162.  
PR 03-AUG-2000; 2000JP-0280988.  
XX  
PA (KYOW ) KYOWA HAKKO KOGYO KK.  
XX  
PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;  
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;  
XX  
WPI: 2001-376931/40.  
DR P-PSDB; AAG90302.  
XX  
PT Novel polynucleotides derived from Coryneform bacteria, for identifying  
PT mutation point of a gene, measuring expression of a gene, analysing  
PT expression profile or pattern of a gene and identifying homologous gene  
XX  
PS Claim 8; SEQ ID NO: 556; 246pp + Sequence Listing; English.  
XX  
CC The present invention provides a number of nucleotide and protein  
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These  
CC are useful for identifying the mutation point of a gene derived from a  
CC mutant of coryneform bacterium, measuring expression amount and  
CC analysing the expression profile or expression pattern of a gene derived  
CC from Coryneform bacterium, and identifying a homolog of a gene derived  
CC from coryneform bacterium. Coryneform bacteria are useful for producing  
CC amino acids, nucleic acids, vitamins, saccharides and organic acids.  
CC particularly L-lysine. The present sequence is a nucleic acid described  
CC in the exemplification of the invention.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from the  
CC European Patent Office.  
XX  
SQ Sequence 2160 BP; 462 A; 569 C; 621 G; 508 T; 0 other;  
Alignment Scores:  
Pred. No.: 286 Length: 2160  
Score: 46.00 Matches: 9  
Percent Similarity: 62.50% Conservative: 6  
Best Local Similarity: 37.50% Mismatches: 9  
Query Match: 35.66% Indels: 0  
DB: 22 Gaps: 0  
US-09-847-539A-6\_COPY\_59\_86 (1-28) x AAH65521 (1-2160)

QY 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLys 23  
:::||||| ||| :::: ||| |||::: |||::: |||::: |||::: |||  
Db 636 ATTGAGCCTTACCGACTGAGGAGGAGGCTTGAAGAGGGCGAGTCCGCAATCTACAAG 577  
QY 24 AlaAspAsnAla 27  
:::|||||  
Db 576 TTCGAGAACGCT 565  
RESULT 28  
AAH81479/c  
ID AAH81479 standard; DNA; 69936 BP.  
XX  
AC AAH81479;  
DT 04-DEC-2000 (first entry)  
XX  
DE N. meningitidis partial DNA sequence gnm\_27 SEQ ID NO: 27.  
XX  
KW Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;  
KW antigen; vaccine; diagnosis; infection; antibacterial; identification;  
KW Meningococcus B; MenB; ds.  
XX  
OS Neisseria meningitidis.  
XX  
PN WO200022430-A2.  
XX  
PD 20-APR-2000.  
XX  
PF 08-OCT-1999; 99WO-US23573.  
XX  
PR 09-OCT-1998; 98US-0103794.  
PR 30-APR-1999; 99US-0132068.  
XX  
PA (CHIR ) CHIRON CORP.  
XX  
PI Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;  
PI Maignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;  
PI Rappuoli R, Pizza M;  
XX  
WPI: 2000-318079/27.  
XX  
PT Isolated nucleotide sequences of Neisseria meningitidis which can be  
PT used in the diagnosis and treatment of N. meningitidis infection and  
PT other Neisserial infections, for example, N.gonorrhoea -  
XX  
PS Claim 7; Page 547-567; 1760pp; English.  
XX  
CC The present invention describes methods of obtaining immunogenic  
CC proteins from Neisseria genomic sequences. AAA81453 to AAA82414  
CC represent specifically claimed Neisseria meningitidis genomic DNA  
CC sequences; AAA81260 to AAA81303 and AAB25620 to AAB25663 represent  
CC Neisseria DNA sequences and their corresponding proteins; AAA81254 to  
CC AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the  
CC isolation of Neisseria meningitidis DNA sequences; and AAA81322 to  
CC AAA81452 represent Neisseria meningitidis MenB polynucleotide ORF  
CC sequences, which are all used in the exemplification of the present  
CC invention. The nucleic acid sequences, protein sequences, and antibodies  
CC against them, can be used in the manufacture of a composition. The  
CC composition can be used as a medicament (or in the manufacture of a  
CC medicament) for treating, preventing or diagnosing infection due to  
CC Neisserial bacteria. For example, some of the identified proteins could  
CC be components of vaccines against Meningococcus B; against all serotypes;  
CC and/or against all pathogenic Neisseriae. Identification of sequences  
CC from the bacterium will also facilitate production of biological probes,  
CC particularly organism-specific probes. Attempts to make efficacious  
CC Meningococcus B vaccines have failed mainly due to antigen tolerance.  
CC Multivalent vaccines have also been tried but none have successfully  
CC overcome antigenic variability. The provision of further, complete  
CC sequences may provide an opportunity to identify secreted or surface  
CC exposed proteins that may be presumed targets for the immune system and  
CC which are not antigenically variable or at least more conserved than  
CC other more variable regions.

AC AAF11343;  
 XX 13-MAR-2001 (first entry)  
 XX Aspergillus niger EST SEQ ID NO:3866.  
 XX  
 XX Multiple gene expression; filamentous fungal cell; EST;  
 KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;  
 KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;  
 KW culture condition; environmental stress; spore morphogenesis;  
 KW metabolic pathway engineering; catabolic pathway engineering; ss.  
 XX  
 OS Aspergillus niger.  
 XX  
 PN WO200056762-A2.  
 XX  
 XX 28-SEP-2000.  
 XX  
 XX 22-MAR-2000; 2000WO-US07781.  
 XX  
 XX 22-MAR-1999; 99US-0273623.  
 XX  
 XX (NOVO ) NOVO NORDISK BIOTECH INC.  
 PA (NOVO ) NOVO NORDISK AS.  
 XX  
 XX Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;  
 XX WPI; 2000-594572/56.  
 DR  
 XX  
 XX Monitoring differential expression of genes in filamentous fungal cells  
 PT uses fluorescence-labeled nucleic acids isolated from the cells and a  
 PT substrate of expressed sequence tags -  
 XX  
 XX Claim 87; Page 1728; 3161pp; English.  
 PS  
 XX  
 CC The present invention describes a method for monitoring differential  
 CC expression of genes in a first filamentous fungal (FF) cell relative to  
 CC expression of the same genes in one or more second filamentous fungal  
 CC cells. The method uses fluorescence-labeled nucleic acids isolated from  
 CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs  
 CC are used in the methods for monitoring differential expression of genes  
 CC in a first filamentous fungal (FF) cell relative to expression of the  
 CC same genes in one or more second filamentous fungal cells. Monitoring  
 CC the global expression of genes from FF cells allows the production  
 CC potential of the microorganisms to be improved. New genes may be  
 CC discovered, possible functions of unknown open reading frames can be  
 CC identified and gene copy number variation and stability can be  
 CC monitored. The expression of genes can be used to study how FF cells  
 CC adapt to changes in culture conditions, environmental stress, spore  
 CC morphogenesis, recombination, metabolic or catabolic pathway  
 CC engineering. Using ESTs provides several advantages over genomic or  
 CC random cDNA clones including elimination of redundancy as one spot on an  
 CC array equals one gene or open reading frame, and organisation of the  
 CC microarrays based on function of the gene products to facilitate  
 CC analysis of the results. AAF07478 to AAF11247 represents ESTs from  
 CC Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus  
 CC niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and  
 CC AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are  
 CC all specifically claimed in the present invention.  
 XX  
 XX Sequence 636 BP; 169 A; 155 C; 161 G; 149 T; 2 other;  
 SQ  
 Alignment Scores:  
 Pred. No.: 61.2 Length: 636  
 Score: 46.00 Matches: 11  
 Percent Similarity: 64.00% Conservative: 5  
 Best Local Similarity: 44.00% Mismatches: 9  
 Query Match: 35.66% Indels: 0  
 DB: 21 Gaps: 0  
 US-09-847-539A-6\_COPY\_59\_86 (1-28) x AAF11343 (1-636)  
 Oy 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaA 20

Db 149 TCTAAAGCTGCAGAGCCTTCCGAGAGATCGCCGCCAGAAATAATGTCAAGTCAGCT 208  
 Oy 21 ValVallytsAlaAsp 25  
 Db 209 GCCTATATAAGCTGAT 223  
 RESULT 26  
 AAH65520  
 ID AAH65520 standard; DNA; 2115 BP.  
 XX  
 AC AAH65520;  
 XX  
 XX 26-SEP-2001 (first entry)  
 XX  
 XX C glutamicum coding sequence fragment SEQ ID NO: 555.  
 DE  
 XX  
 KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;  
 KW organic acid synthesis; ds.  
 XX  
 OS Corynebacterium glutamicum.  
 XX  
 PN EP1108790-A2.  
 XX  
 PD 20-JUN-2001.  
 XX  
 XX 18-DEC-2000; 2000EP-0127688.  
 PF  
 XX 16-DEC-1999; 99JP-0377484.  
 PR  
 PR 07-APR-2000; 2000JP-0159162.  
 PR  
 PR 03-AUG-2000; 2000JP-0280988.  
 XX  
 XX (KYOW ) KYOWA HAKKO KOGYO KK.  
 PA  
 XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;  
 PI Tateishi N, Senoh A, Ikeda M, Ozaki A;  
 PI  
 DR WPI; 2001-376931/40.  
 DR  
 XX P-PSDB; AAG90301.  
 XX  
 PT Novel polynucleotides derived from Coryneform bacteria, for identifying  
 PT mutation point of a gene, measuring expression of a gene, analysing  
 PT expression profile or pattern of a gene and identifying homologous gene  
 PT  
 XX  
 XX Claim 8; SEQ ID NO: 555; 246pp + Sequence Listing; English.  
 PS  
 XX The present invention provides a number of nucleotide and protein  
 CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These  
 CC are useful for identifying the mutation point of a gene derived from a  
 CC mutant of coryneform bacterium, measuring expression amount and  
 CC analysing the expression profile or expression pattern of a gene derived  
 CC from Coryneform bacterium, and identifying a homologue of a gene derived  
 CC from coryneform bacterium. Coryneform bacteria are useful for producing  
 CC amino acids, nucleic acids, vitamins, saccharides and organic acids,  
 CC particularly L-lysine. The present sequence is a nucleic acid described  
 CC in the exemplification of the invention.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from the  
 CC European Patent Office.  
 XX  
 XX Sequence 2115 BP; 492 A; 609 C; 561 G; 453 T; 0 other;  
 SQ  
 Alignment Scores:  
 Pred. No.: 278 Length: 2115  
 Score: 46.00 Matches: 9  
 Percent Similarity: 62.50% Conservative: 6  
 Best Local Similarity: 37.50% Mismatches: 9  
 Query Match: 35.66% Indels: 0  
 DB: 22 Gaps: 0  
 US-09-847-539A-6\_COPY\_59\_86 (1-28) x AAH65520 (1-2115)



```
DB: 23 Gaps: 0
US-09-847-539A-6_COPY_59_86 (1-28) x AAS94455 (1-3711)
Qy 3 AlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVal 22
    |||::: |||
    |||::: |||
Db 1741 GCAGTGATCCAACTGACGCCGACAGCGAGCGTGGCAGTGAAGACGCTCCGGTTGCA 1800
Qy 23 LysAlaAspAsn 26
    |||::: |||
    |||::: |||
Db 1801 CCGCGCGACAAT 1812
RESULT 21
ID ABL25849/c
ABL25849 standard; DNA; 1616 BP.
XX
AC ABL25849;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 29020.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
PI WPI; 2001-656860/75.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 29020; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1616 BP; 457 A; 447 C; 439 G; 273 T; 0 other;
Alignment Scores:
Pred. No.: 88.8 Length: 1616
Score: 48.00 Matches: 12
Percent Similarity: 50.00% Conservative: 1
Best Local Similarity: 46.15% Mismatches: 13
Query Match: 37.21% Indels: 0
DB: 23 Gaps: 0
US-09-847-539A-6_COPY_59_86 (1-28) x ABL25849 (1-1616)
Qy 3 AlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVal 22
    |||::: |||
    |||::: |||
```

```
Db 885 GCGGCTGATGCTGCTGGTGGATGCTGGTGGCGTGGTGGGAGCATGCTGCTGCTGCT 826
    |||::: |||
    |||::: |||
Qy 23 LysAlaAspAsnAlaAla 28
    |||::: |||
    |||::: |||
Db 825 GCTGCTGATGCTGGTTGCA 808
    |||::: |||
    |||::: |||
RESULT 22
ABL25848/c
ID ABL25848 standard; DNA; 4229 BP.
XX
AC ABL25848;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 29017.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
PI WPI; 2001-656860/75.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 29017; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 4229 BP; 1287 A; 1022 C; 928 G; 992 T; 0 other;
Alignment Scores:
Pred. No.: 299 Length: 4229
Score: 48.00 Matches: 12
Percent Similarity: 50.00% Conservative: 1
Best Local Similarity: 46.15% Mismatches: 13
Query Match: 37.21% Indels: 0
DB: 23 Gaps: 0
US-09-847-539A-6_COPY_59_86 (1-28) x ABL25848 (1-4229)
Qy 3 AlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVal 22
    |||::: |||
    |||::: |||
Db 1957 GCGGCTGATGCTGGTGGATGCTGGTGGCGTGGTGGGAGCATGCTGCTGCTGCT 1898
    |||::: |||
    |||::: |||
Qy 23 LysAlaAspAsnAlaAla 28
    |||::: |||
    |||::: |||
```

DB ABA03041 standard; DNA; 2944528 BP.  
ABA03041;  
05-FEB-2002 (first entry)  
Listeria monocytogenes EGD-e genome sequence.  
Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;  
vitamin B12; bacterial infection; disease; ds.  
Listeria monocytogenes.  
WO200177335-A2.  
18-OCT-2001.  
11-APR-2001; 2001WO-FR01118.  
11-APR-2000; 2000FR-0004629.  
(INSP ) INST PASTEUR.  
Buchrieser C, Frangeul L, Couve E, Rusniok C, Fsihi H, Dehoux P;  
Dussurget O, Chetoui F, Nedjari H, Glaser P, Kunst F, Cossart P;  
Daniels J, Goebel W, Kreft J, Kuhn M, Ng E, Vazquez-Boland JA;  
Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;  
Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L;  
Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;  
Madueno E, De Pablos B, Wehland J, Kaerst U, Entian K, Hauf J;  
Rose M, Voss H;  
WPI; 2002-010914/01.  
Genomic sequence for Listeria monocytogenes, useful e.g. for treatment  
and prevention of Listeria and related bacterial infections, and  
related polypeptides -  
Claim 1: SEQ ID No 1; 192pp; French.  
The present sequence is the genome sequence of Listeria monocytogenes  
EGD-e. This sequence and fragments of this sequence are useful for  
selecting probes and primers for detecting genes in L. monocytogenes and  
related organisms, and to study genetic polymorphisms and other genomes.  
Proteins (ABBA47297-ABB50149) expressed from the present sequence are  
useful for raising specific antibodies, identification of L.  
monocytogenes and related organisms, and for biosynthesis and  
biodegradation, especially biosynthesis of Vitamin B12. This sequence and  
proteins encoded by it are also useful for selecting compounds that  
regulate gene expression and cell replication and modulate L.  
monocytogenes-related diseases. In addition, this sequence and proteins  
encoded by it are useful in pharmaceutical and vaccines compositions for  
the treatment or prevention of infections by L. monocytogenes and related  
organisms.  
Note: The sequence data for this patent did not form part of the printed  
specification, but was obtained in electronic format directly from WIPO  
at ftp.wipo.int/pub/published\_pct\_sequences.  
SQ Sequence 2944528 BP; 914202 A; 563301 C; 555061 G; 911964 T; 0 other;  
Alignment Scores:  
Pred. No.: 3.25e+05 Length: 2944528  
Score: 51.00 Matches: 10  
Percent Similarity: 65.38% Conservative: 7  
Best Local Similarity: 38.46% Mismatches: 9  
Query Match: 39.53% Indels: 0  
DB: 24 Gaps: 0  
US-09-847-539A-6\_COPY\_59\_86 (1-28) x ABA03041 (1-2944528)  
QY 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaLa 20  
: : : : : ||| ||| : : : : : ||| ||| : : : : : ||| ||| : : : : : ||| |||  
Db 51231 GCTGAACAAAGAACAACTAAGTGTGATGAACAGATCGACTTGACCGAGAGATGCAGCT 51290

QY 21 ValValLysAlaAspAsn 26  
||| : : : |||  
Db 51291 GGAGTTGAAAAGAAAAC 51308  
RESULT 20  
AAS94455  
ID AAS94455 standard; CDNA; 3711 BP.  
XX AAS94455;  
AC AAS94455;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE DNA encoding novel human diagnostic protein #30259.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DR WPI; 2001-639362/73.  
DR P-PSDB; ABG30268.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX  
PS Claim 1; SEQ ID No 30259; 103pp; English.  
XX  
CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 3711 BP; 848 A; 945 C; 1056 G; 862 T; 0 other;  
Alignment Scores:  
Pred. No.: 114 Length: 3711  
Score: 50.00 Matches: 11  
Percent Similarity: 66.67% Conservative: 5  
Best Local Similarity: 45.83% Mismatches: 8  
Query Match: 38.76% Indels: 0



```

Db 228 GATCGGCGGAATCATATAGAGACACACAGCAGCAGAGGAGGAGGAACAAGCGCG 169
    |||
Qy 22 VallysAlaAspAsnAla 27
    |||::: ||| ||||| ||| ||||| |||
Db 168 GTACGAGGAGATCGTGCC 151
    |||::: ||| |||

RESULT 17
ID ABL27360
ABL27360 standard; DNA; 2618 BP.
XX ABL27360;
XX
Dt 26-MAR-2002 (first entry)
XX
De Drosophila melanogaster genomic polynucleotide SEQ ID NO 33553.
XX
Kw Drosophila; developmental biology; cell signalling; insecticide;
Kw pharmaceutical; gene; ds.
XX
Os Drosophila melanogaster.
XX
Pn WO200171042-A2.
XX
Pd 27-SEP-2001.
XX
Pf 23-MAR-2001; 2001WO-US09231.
XX
Pr 23-MAR-2000; 2000US-191637P.
Pr 11-JUL-2000; 2000US-0614150.
XX
Pa (PEKE ) PE CORP NY.
XX
Pi Venter JC, Adams M, Li PWD, Myers EW;
XX
Dr WPI; 2001-656860/75.
XX
Pt New isolated nucleic acid detection reagent for detecting 1000 or more
Pt genes from Drosophila and for elucidating cell signalling and cell-cell
Pt interactions -
XX
Ps Claim 1; SEQ ID NO 33553; 2lpp + Sequence Listing; English.
XX
Cc The invention relates to an isolated nucleic acid detection reagent
Cc capable of detecting 1000 or more genes from Drosophila. The invention is
Cc useful in developmental biology and in elucidating cell signalling and
Cc cell-cell interactions in higher eukaryotes for the development of
Cc insecticides, therapeutics and pharmaceutical drugs. The invention
Cc discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
Cc sequences (ABL01840-ABL16175) and the encoded proteins
Cc (ABB57737-ABB72072).
Cc The sequence data for this patent did not form part of the printed
Cc specification, but was obtained in electronic format directly from WIPO
Cc at ftp.wipo.int/pub/published_pct_sequences.
XX
Sq Sequence 2618 BP; 669 A; 603 C; 652 G; 694 T; 0 other;

Alignment Scores:
Pred. No.: 49 Length: 2618
Score: 51.00 Matches: 11
Percent Similarity: 53.85% Conservative: 3
Best Local Similarity: 42.31% Mismatches: 12
Query Match: 39.53% Indels: 0
DB: 23 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x ABL27360 (1-2618)
Qy 2 AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaVal 21
    |||
    |||::: ||| ||||| ||| ||||| |||
Db 467 GATCGGCGGAATCATATAGAGACACACAGCAGCAGAGGAGGAGGAACAAGCGCG 526
    |||
Qy 22 VallysAlaAspAsnAla 27
    |||::: ||| |||

" . . .

```

```

Db 527 GTACGAGGAGATCGTGCC 544
RESULT 18
ABL12682
ID ABL12682 standard; cDNA; 4221 BP.
XX
Ac ABL12682;
XX
Dt 26-MAR-2002 (first entry)
XX
De Drosophila melanogaster expressed polynucleotide SEQ ID NO 32528.
XX
Kw Drosophila; developmental biology; cell signalling; insecticide;
Kw pharmaceutical; gene; ss.
XX
Os Drosophila melanogaster.
XX
Pn WO200171042-A2.
XX
Pd 27-SEP-2001.
XX
Pf 23-MAR-2001; 2001WO-US09231.
XX
Pr 23-MAR-2000; 2000US-191637P.
Pr 11-JUL-2000; 2000US-0614150.
XX
Pa (PEKE ) PE CORP NY.
XX
Pi Venter JC, Adams M, Li PWD, Myers EW;
XX
Dr WPI; 2001-656860/75.
Dr P-PSDB; ABB68579.
XX
Pt New isolated nucleic acid detection reagent for detecting 1000 or more
Pt genes from Drosophila and for elucidating cell signalling and cell-cell
Pt interactions -
XX
Ps Claim 1; SEQ ID NO 32528; 2lpp + Sequence Listing; English.
XX
Cc The invention relates to an isolated nucleic acid detection reagent
Cc capable of detecting 1000 or more genes from Drosophila. The invention is
Cc useful in developmental biology and in elucidating cell signalling and
Cc cell-cell interactions in higher eukaryotes for the development of
Cc insecticides, therapeutics and pharmaceutical drugs. The invention
Cc discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
Cc sequences (ABL01840-ABL16175) and the encoded proteins
Cc (ABB57737-ABB72072).
Cc The sequence data for this patent did not form part of the printed
Cc specification, but was obtained in electronic format directly from WIPO
Cc at ftp.wipo.int/pub/published_pct_sequences.
XX
Sq Sequence 4221 BP; 1022 A; 1036 C; 1072 G; 1091 T; 0 other;

Alignment Scores:
Pred. No.: 89.4 Length: 4221
Score: 51.00 Matches: 11
Percent Similarity: 53.85% Conservative: 3
Best Local Similarity: 42.31% Mismatches: 12
Query Match: 39.53% Indels: 0
DB: 23 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x ABL12682 (1-4221)
Qy 2 AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaVal 21
    |||
    |||::: ||| ||||| ||| ||||| |||
Db 2994 GATCGGCGGAATCATATAGAGACACACAGCAGCAGAGGAGGAGGAACAAGCGCG 3053
    |||
Qy 22 VallysAlaAspAsnAla 27
    |||::: ||| |||
Db 3054 GTACGAGGAGATCGTGCC 3071
    |||

RESULT 19
ABA03041

```

Score: 52.00 Matches: 11  
Percent Similarity: 65.22% Conservative: 4  
Best Local Similarity: 47.83% Mismatches: 8  
Query Match: 40.31% Indels: 0  
DB: 23 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x ABL25211 (1-16135)

QY 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVallys 23  
|||||  
Db 5971 CTGGAAGCTCTGTTCAGCAGGTGGAGTCTCTCCAAAGCGCGCAGTTGCTCTCTGTGGCA 5912  
|||||

QY 24 AlaAspAsn 26  
|||||

Db 5911 GCCGACAAT 5903  
|||||

#### RESULT 15

ABL25210  
ID ABL25210 standard; DNA: 18603 BP.

XX AC ABL25210;

XX DT 26-MAR-2002 (first entry)

XX DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 27103.

XX KW Drosophila; developmental biology; cell signalling; insecticide;

XX KW pharmaceutical; gene; ds.

XX OS Drosophila melanogaster.

XX PN WO200171042-A2.

XX PD 27-SEP-2001.

XX PF 23-MAR-2001; 2001WO-US09231.

XX PR 23-MAR-2000; 2000US-191637P.

XX PR 11-JUL-2000; 2000US-0614150.

XX PA (PEKE ) PE CORP NY.

XX PI Venter JC, Adams M, Li PWD, Myers EW;

XX DR WPI; 2001-656860/75.

XX PT New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -

XX PS Claim 1; SEQ ID NO 27103; 21pp + Sequence Listing; English.

XX CC The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-ABB72072).

XX CC The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 18603 BP; 4763 A; 4409 C; 4291 G; 5140 T; 0 other;

#### Alignment Scores:

Pred. No.: 388 Length: 18603  
Score: 52.00 Matches: 11  
Percent Similarity: 65.22% Conservative: 4  
Best Local Similarity: 47.83% Mismatches: 8  
Query Match: 40.31% Indels: 0  
DB: 23 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x ABL25210 (1-186603)

QY 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVallys 23  
|||||  
Db 11563 CTGGAAGCTCTGTTCAGCAGGTGGAGTCTCTCCAAAGCGCGCAGTTGCTCTCTGTGGCA 11622  
|||||

QY 24 AlaAspAsn 26  
|||||

Db 11623 GCCGACAAT 11631  
|||||

#### RESULT 16

ABL12683/C  
ID ABL12683 standard; cDNA: 1960 BP.

XX AC ABL12683;

XX DT 26-MAR-2002 (first entry)

XX DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 32531.

XX KW Drosophila; developmental biology; cell signalling; insecticide;

XX KW pharmaceutical; gene; ss.

XX OS Drosophila melanogaster.

XX PN WO200171042-A2.

XX PD 27-SEP-2001.

XX PF 23-MAR-2001; 2001WO-US09231.

XX PR 23-MAR-2000; 2000US-191637P.

XX PR 11-JUL-2000; 2000US-0614150.

XX PA (PEKE ) PE CORP NY.

XX PI Venter JC, Adams M, Li PWD, Myers EW;

XX DR WPI; 2001-656860/75.

XX DR P-PSDB; ABB68580.

XX PT New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -

XX PS Claim 1; SEQ ID NO 32531; 21pp + Sequence Listing; English.

XX CC The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-ABB72072).

XX CC The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 1960 BP; 505 A; 526 C; 485 G; 444 T; 0 other;

#### Alignment Scores:

Pred. No.: 34 Length: 1960  
Score: 51.00 Matches: 11  
Percent Similarity: 53.85% Conservative: 3  
Best Local Similarity: 42.31% Mismatches: 12  
Query Match: 39.53% Indels: 0  
DB: 23 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x ABL12683 (1-1960)

QY 2 AspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaVal 21

PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157755.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0158393.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.

PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.  
  
Alignment Scores:  
Pred. No.: 12.7 Length: 1230  
Score: 52.00 Matches: 11  
Percent Similarity: 72.73% Conservative: 5  
Best Local Similarity: 50.00% Mismatches: 6  
Query Match: 40.31% Indels: 0  
DB: 21 Gaps: 0  
  
US-09-847-539A-6\_COPY\_59\_86 (1-28) x AAC33266 (1-1230)  
  
Qy 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluAlaAlaValVallys 23  
||| |||||:::||||: ||| ||| :::::|||||::: |||  
Db 518 CTTCCCGCAGCTCTCCGACAAAAGGACATGCTCAAAATCAAAAAGCTCAATGCAAAAG 459  
  
Qy 24 Alaasp 25  
|||||  
Db 458 CCGGAT 453  
  
RESULT 14  
ABL25211/C  
ID ABL25211 standard; DNA; 16135 BP.  
XX AC ABL25211;  
XX DT 26-MAR-2002 (first entry)  
XX DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 27106.  
XX KW Drosophila; developmental biology; cell signalling; insecticide;  
XX KW pharmaceutical; gene; ds.  
XX OS Drosophila melanogaster.  
XX PN WO200171042-A2.  
XX PD 27-SEP-2001.  
XX PF 23-MAR-2001; 2001WO-US09231.  
XX PF 23-MAR-2000; 2000US-191637P.  
XX PR 11-JUL-2000; 2000US-0614150.  
XX PA (PEKE ) PE CORP NY.  
XX PI Venter JC, Adams M, Li PWD, Myers EW;  
XX WPI; 2001-656860/75.  
XX  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
genes from Drosophila and for elucidating cell signalling and cell-cell  
interactions -  
XX  
XX Claim 1; SEQ ID NO 27106; 2lpp + Sequence Listing; English.  
XX  
XX The invention relates to an isolated nucleic acid detection reagent  
capable of detecting 1000 or more genes from Drosophila. The invention is  
useful in developmental biology and in elucidating cell signalling and  
cell-cell interactions in higher eukaryotes for the development of  
insecticides, therapeutics and pharmaceutical drugs. The invention  
discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
sequences (ABL01840-ABL16175) and the encoded proteins  
CC (ABB57737-ABB72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 16135 BP; 4290 A; 3936 C; 4034 G; 3875 T; 0 other;  
  
Alignment Scores: 324 Length: 16135  
Pred. No.: 324

Alignment Scores: 9.78 Length: 1002  
Pred. No.: 52.00 Matches: 11  
Score: 72.73% Mismatches: 5  
Best Local Similarity: 50.00% Indels: 0  
Query Match: 40.31% Gaps: 0  
DB: 21

US-09-847-539a-6\_COPY\_59\_86 (1-28) x AAC54228 (1-1002)

QY 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValValLys 23  
||| |||||:||||: ||| ||| :|||:||||: |||  
Db 442 CTTCCGCGACTCTCCGACAAAGGACATCCCTCAATTCAAAAGCTGCATCGCAAG 383

QY 24 AlaAsp 25

|||||  
Db 382 GCGGAT 377

RESULT 13

AAC33266/C

ID AAC33266 standard; DNA; 1230 BP.

XX AC AAC33266;

XX DT 17-OCT-2000 (first entry)

XX DE Arabidopsis thaliana DNA fragment SEQ ID NO: 2396.

XX KW Hybridisation assay; genetic mapping; gene expression control;  
KW protein identification; signal transduction pathway;  
KW metabolic pathway; promoter; termination sequence; ss.

XX OS Arabidopsis thaliana.

XX PN EP1033405-A2.

XX PD 06-SEP-2000.

XX PF 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

PR 09-MAR-1999; 99US-0123180.

PR 23-MAR-1999; 99US-0123548.

PR 25-MAR-1999; 99US-0123788.

PR 29-MAR-1999; 99US-0126264.

PR 01-APR-1999; 99US-0126785.

PR 06-APR-1999; 99US-0127462.

PR 08-APR-1999; 99US-0128234.

PR 16-APR-1999; 99US-0128714.

PR 19-APR-1999; 99US-0129845.

PR 21-APR-1999; 99US-0130077.

PR 23-APR-1999; 99US-0130449.

PR 28-APR-1999; 99US-0130510.

PR 28-APR-1999; 99US-0130891.

PR 30-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 04-MAY-1999; 99US-0132407.

PR 05-MAY-1999; 99US-0132484.

PR 06-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

PR 07-MAY-1999; 99US-0132487.

PR 11-MAY-1999; 99US-0132863.

PR 14-MAY-1999; 99US-0134256.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134221.

PR 14-MAY-1999; 99US-0134370.

PR 18-MAY-1999; 99US-0134768.

PR 19-MAY-1999; 99US-0134941.

PR 20-MAY-1999; 99US-0135124.

PR 21-MAY-1999; 99US-0135353.

PR 24-MAY-1999; 99US-0135629.

PR 25-MAY-1999; 99US-0136021.

PR 27-MAY-1999; 99US-0136392.

PR 28-MAY-1999; 99US-0136782.

PR 01-JUN-1999; 99US-0137222.

PR 03-JUN-1999; 99US-0137528.

PR 04-JUN-1999; 99US-0137502.

PR 07-JUN-1999; 99US-0137724.

PR 08-JUN-1999; 99US-0138094.

PR 10-JUN-1999; 99US-0138540.

PR 10-JUN-1999; 99US-0138847.

PR 14-JUN-1999; 99US-0139119.

PR 16-JUN-1999; 99US-0139452.

PR 16-JUN-1999; 99US-0139453.

PR 17-JUN-1999; 99US-0139452.

PR 18-JUN-1999; 99US-0139454.

PR 18-JUN-1999; 99US-0139455.

PR 18-JUN-1999; 99US-0139456.

PR 18-JUN-1999; 99US-0139457.

PR 18-JUN-1999; 99US-0139458.

PR 18-JUN-1999; 99US-0139459.

PR 18-JUN-1999; 99US-0139460.

PR 18-JUN-1999; 99US-0139461.

PR 18-JUN-1999; 99US-0139462.

PR 18-JUN-1999; 99US-0139463.

PR 18-JUN-1999; 99US-0139750.

PR 18-JUN-1999; 99US-0139763.

PR 21-JUN-1999; 99US-0139817.

PR 22-JUN-1999; 99US-0139899.

PR 23-JUN-1999; 99US-0140353.

PR 23-JUN-1999; 99US-0140354.

PR 24-JUN-1999; 99US-0140695.

PR 28-JUN-1999; 99US-0140823.

PR 29-JUN-1999; 99US-0140991.

PR 30-JUN-1999; 99US-0141287.

PR 01-JUL-1999; 99US-0141842.

PR 02-JUL-1999; 99US-0142055.

PR 06-JUL-1999; 99US-0142390.

PR 08-JUL-1999; 99US-0142803.

PR 09-JUL-1999; 99US-0142920.

PR 12-JUL-1999; 99US-0142977.

PR 13-JUL-1999; 99US-0143542.

PR 14-JUL-1999; 99US-0143624.

PR 15-JUL-1999; 99US-0144005.

PR 16-JUL-1999; 99US-0144085.

PR 16-JUL-1999; 99US-0144086.

PR 19-JUL-1999; 99US-0144325.

PR 19-JUL-1999; 99US-0144331.

PR 19-JUL-1999; 99US-0144332.

PR 19-JUL-1999; 99US-0144333.

PR 19-JUL-1999; 99US-0144334.

PR 19-JUL-1999; 99US-0144335.

PR 20-JUL-1999; 99US-0144352.

PR 20-JUL-1999; 99US-0144632.

PR 20-JUL-1999; 99US-0144884.

PR 21-JUL-1999; 99US-0144814.

PR 21-JUL-1999; 99US-0145086.

PR 21-JUL-1999; 99US-0145088.

PR 22-JUL-1999; 99US-0145085.

PR 22-JUL-1999; 99US-0145087.

PR 22-JUL-1999; 99US-0145089.

PR 22-JUL-1999; 99US-0145192.

PR 23-JUL-1999; 99US-0145145.

PR 23-JUL-1999; 99US-0145218.

PR 23-JUL-1999; 99US-0145224.

PR 26-JUL-1999; 99US-0145276.

PR 27-JUL-1999; 99US-0145913.

PR 27-JUL-1999; 99US-0145918.

PR 27-JUL-1999; 99US-0145919.

PR 28-JUL-1999; 99US-0145951.

PR 02-AUG-1999; 99US-0146386.

PR 02-AUG-1999; 99US-0146388.

PR 02-AUG-1999; 99US-0146389.

PR	28-MAY-1999;	99US-0136782.	PR	04-AUG-1999;	99US-0147302.
PR	01-JUN-1999;	99US-0137222.	PR	05-AUG-1999;	99US-0147192.
PR	03-JUN-1999;	99US-0137528.	PR	05-AUG-1999;	99US-0147260.
PR	04-JUN-1999;	99US-0137502.	PR	06-AUG-1999;	99US-0147303.
PR	07-JUN-1999;	99US-0137724.	PR	06-AUG-1999;	99US-0147416.
PR	08-JUN-1999;	99US-0138094.	PR	09-AUG-1999;	99US-0147493.
PR	10-JUN-1999;	99US-0138540.	PR	09-AUG-1999;	99US-0147935.
PR	10-JUN-1999;	99US-0138847.	PR	10-AUG-1999;	99US-0148171.
PR	14-JUN-1999;	99US-0139119.	PR	11-AUG-1999;	99US-0148319.
PR	16-JUN-1999;	99US-0139452.	PR	12-AUG-1999;	99US-0148341.
PR	16-JUN-1999;	99US-0139453.	PR	13-AUG-1999;	99US-0148565.
PR	17-JUN-1999;	99US-0139492.	PR	13-AUG-1999;	99US-0148684.
PR	18-JUN-1999;	99US-0139434.	PR	16-AUG-1999;	99US-0149368.
PR	18-JUN-1999;	99US-0139455.	PR	17-AUG-1999;	99US-0149175.
PR	18-JUN-1999;	99US-0139456.	PR	18-AUG-1999;	99US-0149426.
PR	18-JUN-1999;	99US-0139457.	PR	20-AUG-1999;	99US-0149722.
PR	18-JUN-1999;	99US-0139458.	PR	20-AUG-1999;	99US-0149723.
PR	18-JUN-1999;	99US-0139459.	PR	20-AUG-1999;	99US-0149929.
PR	18-JUN-1999;	99US-0139460.	PR	23-AUG-1999;	99US-0149902.
PR	18-JUN-1999;	99US-0139461.	PR	23-AUG-1999;	99US-0149930.
PR	18-JUN-1999;	99US-0139462.	PR	25-AUG-1999;	99US-0150566.
PR	18-JUN-1999;	99US-0139463.	PR	26-AUG-1999;	99US-0150884.
PR	18-JUN-1999;	99US-0139750.	PR	27-AUG-1999;	99US-0151065.
PR	18-JUN-1999;	99US-0139763.	PR	27-AUG-1999;	99US-0151066.
PR	21-JUN-1999;	99US-0139817.	PR	27-AUG-1999;	99US-0151080.
PR	22-JUN-1999;	99US-0139899.	PR	30-AUG-1999;	99US-0151303.
PR	23-JUN-1999;	99US-0140353.	PR	31-AUG-1999;	99US-0151438.
PR	23-JUN-1999;	99US-0140354.	PR	01-SEP-1999;	99US-0151930.
PR	24-JUN-1999;	99US-0140695.	PR	07-SEP-1999;	99US-0152363.
PR	28-JUN-1999;	99US-0140823.	PR	10-SEP-1999;	99US-0153070.
PR	29-JUN-1999;	99US-0140991.	PR	13-SEP-1999;	99US-0153758.
PR	30-JUN-1999;	99US-0141287.	PR	15-SEP-1999;	99US-0154018.
PR	01-JUL-1999;	99US-0141842.	PR	16-SEP-1999;	99US-0154039.
PR	01-JUL-1999;	99US-0142154.	PR	20-SEP-1999;	99US-0154779.
PR	02-JUL-1999;	99US-0142055.	PR	22-SEP-1999;	99US-0155139.
PR	06-JUL-1999;	99US-0142390.	PR	23-SEP-1999;	99US-0155486.
PR	08-JUL-1999;	99US-0142803.	PR	24-SEP-1999;	99US-0155659.
PR	09-JUL-1999;	99US-0142920.	PR	28-SEP-1999;	99US-0156458.
PR	12-JUL-1999;	99US-0142977.	PR	29-SEP-1999;	99US-0156596.
PR	13-JUL-1999;	99US-0143542.	PR	04-OCT-1999;	99US-0157117.
PR	14-JUL-1999;	99US-0143624.	PR	05-OCT-1999;	99US-0157753.
PR	15-JUL-1999;	99US-0144005.	PR	06-OCT-1999;	99US-0157865.
PR	16-JUL-1999;	99US-0144085.	PR	07-OCT-1999;	99US-0158029.
PR	16-JUL-1999;	99US-0144086.	PR	08-OCT-1999;	99US-0158232.
PR	19-JUL-1999;	99US-0144325.	PR	12-OCT-1999;	99US-0158369.
PR	19-JUL-1999;	99US-0144331.	PR	13-OCT-1999;	99US-0159293.
PR	19-JUL-1999;	99US-0144332.	PR	13-OCT-1999;	99US-0159294.
PR	19-JUL-1999;	99US-0144333.	PR	13-OCT-1999;	99US-0159295.
PR	19-JUL-1999;	99US-0144334.	PR	14-OCT-1999;	99US-0159329.
PR	19-JUL-1999;	99US-0144335.	PR	14-OCT-1999;	99US-0159330.
PR	20-JUL-1999;	99US-0144352.	PR	14-OCT-1999;	99US-0159331.
PR	20-JUL-1999;	99US-0144632.	PR	14-OCT-1999;	99US-0159637.
PR	20-JUL-1999;	99US-0144884.	PR	14-OCT-1999;	99US-0159638.
PR	21-JUL-1999;	99US-0144814.	PR	18-OCT-1999;	99US-0159584.
PR	21-JUL-1999;	99US-0145086.	PR	21-OCT-1999;	99US-0160741.
PR	21-JUL-1999;	99US-0145088.	PR	21-OCT-1999;	99US-0160767.
PR	22-JUL-1999;	99US-0145085.	PR	21-OCT-1999;	99US-0160768.
PR	22-JUL-1999;	99US-0145087.	PR	21-OCT-1999;	99US-0160770.
PR	22-JUL-1999;	99US-0145089.	PR	21-OCT-1999;	99US-0160814.
PR	22-JUL-1999;	99US-0145192.	PR	21-OCT-1999;	99US-0160815.
PR	23-JUL-1999;	99US-0145145.	PR	22-OCT-1999;	99US-0160980.
PR	23-JUL-1999;	99US-0145218.	PR	22-OCT-1999;	99US-0160981.
PR	23-JUL-1999;	99US-0145224.	PR	22-OCT-1999;	99US-0160989.
PR	26-JUL-1999;	99US-0145276.	PR	25-OCT-1999;	99US-0161404.
PR	27-JUL-1999;	99US-0145913.	PR	25-OCT-1999;	99US-0161405.
PR	27-JUL-1999;	99US-0145918.	PR	25-OCT-1999;	99US-0161406.
PR	27-JUL-1999;	99US-0145919.	PR	26-OCT-1999;	99US-0161359.
PR	28-JUL-1999;	99US-0145951.	PR	26-OCT-1999;	99US-0161360.
PR	02-AUG-1999;	99US-0146386.	PR	26-OCT-1999;	99US-0161361.
PR	02-AUG-1999;	99US-0146388.	PR	28-OCT-1999;	99US-0161920.
PR	02-AUG-1999;	99US-0146389.	PR	28-OCT-1999;	99US-0161992.
PR	03-AUG-1999;	99US-0147038.	PR	28-OCT-1999;	99US-0161993.
PR	04-AUG-1999;	99US-0147204.	PR	29-OCT-1999;	99US-0162142.

```
XX
SQ Sequence 3349 BP; 808 A; 953 C; 851 G; 737 T; 0 other;

Alignment Scores:
Pred. No.: 20.1 Length: 3349
Score: 54.00 Matches: 12
Percent Similarity: 66.67% Conservative: 4
Best Local Similarity: 50.00% Mismatches: 8
Query Match: 41.86% Indels: 0
DB: 23 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x ABL13527 (1-3349)
QY 5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluAlaAlaValValLysAla 24
Db 1122 GAGGAGGAAGCTGACGAGGAGGAGCGCGCGGAGGAGGATCCGCGCGGCTGCCGCC 1063
QY 25 AspAsnAlaAla 28
Db 1062 GATGCTGCGGCG 1051

RESULT 11
ABL13526/C
ID ABL13526 standard; cDNA; 6398 BP.
XX
AC ABL13526;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 35060.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
WPI: 2001-656860/75.
DR P-PSDB; ABB69423.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 35060; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABB57737-ABB72072),
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 6398 BP; 1688 A; 1632 C; 1352 G; 1726 T; 0 other;

Alignment Scores:
```

```

Pred. No.: 45.3 Length: 6398
Score: 54.00 Matches: 12
Percent Similarity: 66.67% Conservative: 4
Best Local Similarity: 50.00% Mismatches: 8
Query Match: 41.86% Indels: 0
DB: 23 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x ABL13526 (1-6398)
QY 5 GluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluAlaAlaValValLysAla 24
Db 2298 GAGGAGGAAGCTGACGAGGAGGAGCGCGCGGAGGAGGATCCGCGCGGCTGCCGCC 2239
QY 25 AspAsnAlaAla 28
Db 2238 GATGCTGCGGCG 2227

RESULT 12
AAC54228/C
ID AAC54228 standard; DNA; 1002 BP.
XX
AC AAC54228;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 77144.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
```

CC and the predicted coding regions defined by computer evaluation. To  
CC identify likely *H. pylori* antigens for vaccine development, the amino  
CC acid sequences predicted from various ORF were analysed for significant  
CC homology to other known or exported membrane proteins. Having identified  
CC and determined the sequences of interest, particular regions can be  
CC isolated from *H. pylori* by PCR amplification for recombinant polypeptide  
CC production, e.g. in *E. coli* hosts.

XX  
SQ Sequence 666 BP; 250 A; 115 C; 119 G; 182 T; 0 other;

## Alignment Scores:

Pred. No.: 0.787 Length: 666  
Score: 57.00 Matches: 12  
Percent Similarity: 69.57% Conservatve: 4  
Best Local Similarity: 52.17% Mismatches: 7  
Query Match: 44.19% Indels: 0  
DB: 18 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x AAT67891 (1-666)

QY 4 LeuGluaLeuAlaAspGlnThrAspAlaLeuGlnSerGlucLuAlaAlaValVallys 23  
DB 229 ATACAGCCCTACAGGAGCAATTTGACGCTTTAGATTCTCAAGAAAAAGTCGTAGCATA 288  
|||||

QY 24 AlaAspAsn 26

DB 289 TGGGATAAC 297  
|||||

## RESULT 9

AAX30461

ID AAX30461 standard; DNA; 1239 BP.

XX

AC AAX30461;

XX

DT 08-JUN-1999 (first entry)

XX

DE H. pylori secreted protein ORF 09cel0413\_35336707\_f2\_9.

XX

XX Vaccine; probe; diagnostic; ORF; cell envelope protein;

KW secreted protein; cellular protein; ds.

XX

OS Helicobacter pylori.

XX

PN WO9818323-A1.

XX

PD 07-MAY-1998.

XX

PF 28-OCT-1997; 97WO-US19575.

XX

PR 14-JUL-1997; 97US-0891928.

PR

PR 28-OCT-1996; 96US-0739150.

PR

PR 06-DEC-1996; 96US-0759739.

XX

XX (ASTR ) ASTRA AB.

PA

PI Alm RA, Smith D;

XX

DR WPI; 1998-271811/24.

XX

DR P-PSDB; AAY10994.

XX

XX Helicobacter pylori nucleic acids and proteins - used to develop

PT products for the detection, prevention and treatment of *H. pylori*

PT infections

XX

PS Claims 3, 4; Page 126-127; 279pp; English.

XX

XX Recombinant or substantially pure preparations of *H. pylori* polypeptides

CC are disclosed, together with the nucleic acids encoding them. In all,

CC 73 ORFs are shown. The proteins are variously cell envelope proteins,

CC secreted proteins or other cellular proteins. Vaccines containing at least

CC nucleic acids or proteins are claimed, as are probes containing at least

CC 8 nucleotides from the nucleic acid sequences. The vaccines are useful

CC for treating or reducing the risk of *H. pylori* infections, and the

CC probes can be used diagnostically for detecting the presence of  
CC Helicobacter in a sample. The products are also of use in screening  
CC for compounds having the ability to interfere with the *H. pylori* life  
CC cycle or to inhibit *H. pylori* infection.

XX  
SQ Sequence 1239 BP; 472 A; 234 C; 206 G; 327 T; 0 other;

## Alignment Scores:

Pred. No.: 1.72 Length: 1239  
Score: 57.00 Matches: 12  
Percent Similarity: 69.57% Conservatve: 4  
Best Local Similarity: 52.17% Mismatches: 7  
Query Match: 44.19% Indels: 0  
DB: 19 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x AAX30461 (1-1239)

QY 4 LeuGluaLeuAlaAspGlnThrAspAlaLeuGlnSerGlucLuAlaAlaValVallys 23  
DB 145 ATACAAGCCCTACAGGAGCAATTTGACGCTTTAGATTCTCAAGAAAAAGTCGTAGCATA 204  
|||||

QY 24 AlaAspAsn 26

DB 205 TGGGATAAC 213  
|||||

## RESULT 10

ABL13527/C

ID ABL13527 standard; cDNA; 3349 BP.

XX

AC ABL13527;

XX

DT 26-MAR-2002 (first entry)

XX

DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 35063.

XX

XX Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ss.

XX

OS Drosophila melanogaster.

XX

PN WO200171042-A2.

XX

PD 27-SEP-2001.

XX

PF 23-MAR-2001; 2001WO-US09231.

XX

PR 23-MAR-2000; 2000US-191637P.

PR

PR 11-JUL-2000; 2000US-0614150.

XX

XX (PEKE ) PE CORP NY.

XX

XX Venter JC, Adams M, Li PWD, Myers EW;

XX

XX WPI; 2001-656860/75.

DR

DR P-PSDB; ABB69424.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Qy 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAla 20  
Db 556 AGTGACGCTTAGAAGCATGCGGATCAACAGACGCTTTACATCAAGAGAGCTGCG 615  
Qy 21 ValValLysAlaAspAsnAlaAla 28  
Db 616 GTTGTTAAAGCGGATACGCTGCT 639  
RESULT 7  
AAT67618  
ID AAT67618 standard; DNA: 576 BP.  
XX  
AC AAT67618;  
XX  
XX 10-JUL-1997 (first entry)  
XX H. pylori secreted or periplasmic protein ORF 35336707.aa.  
XX  
XX Cytoplasmic; vaccine; prevention; treatment; infection; envelope;  
KW identification; binding compound; bacterium; life cycle; activator;  
KW bacteria; inhibitor; duodenal ulcer disease; chronic gastritis;  
KW diagnosis; ds.  
XX  
OS Helicobacter pylori.  
XX  
XX Key Location/Qualifiers  
FH 1..576  
FT /\*tag= a  
FT /trans\_except= (pos: 541..543, aa: Glu)  
FT /trans\_except= (pos: 547..549, aa: Asp)  
FT /note= "no stop codon given"  
FT 559..561  
FT /\*tag= b  
FT /note= "encodes Asn"  
FT 574..576  
FT /\*tag= c  
FT /note= "encodes Asn"  
FT 577..578  
FT /\*tag= d  
FT /note= "encodes Asn"  
XX  
XX W09640893-A1.  
XX 19-DEC-1996.  
XX 06-JUN-1996; 96WO-US09122.  
XX 01-APR-1996; 96US-0630405.  
XX 07-JUN-1995; 95US-0487032.  
XX (ASTR ) ASTRA AB.  
XX Berglindh OT, Smith D, Mellgaard BL;  
XX WPI; 1997-052306/05.  
XX P-PSDB; AAW20445.  
XX  
XX Helicobacter pylori nucleic acid sequences and related  
PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori  
PT infection, and to detect Helicobacter  
XX  
XX Claim 23; Page 268; 1481pp; English.  
XX  
XX This sequence encodes a H. pylori secreted or periplasmic protein.  
CC The protein may be used in a vaccine to prevent or treat H. pylori  
CC infection or to identify H. pylori polypeptide binding compounds,  
CC useful as potential H. pylori life cycle activators or inhibitors.  
CC The genomic sequence of H. pylori (ATCC 55679) was determined from  
CC overlapping contigs generated by mechanically shearing the bacterial  
CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,  
CC and the predicted coding regions defined by computer evaluation. To  
CC identify likely H. pylori antigens for vaccine development, the amino  
CC acid sequences predicted from various ORF were analysed for significant  
CC homology to other known or exported membrane proteins. Having identified  
CC and determined the sequences of interest, particular regions can be  
CC isolated from H. pylori by PCR amplification for recombinant polypeptide

CC production, e.g. in E. coli hosts.  
XX  
SQ Sequence 576 BP; 217 A; 97 C; 95 G; 161 T; 6 other;  
Alignment Scores:  
Pred. No.: 0.655 Length: 576  
Score: 57.00 Matches: 12  
Percent Similarity: 69.57% Conservative: 4  
Best Local Similarity: 52.17% Mismatches: 7  
Query Match: 44.19% Indels: 0  
DB: 18 Gaps: 0  
US-09-847-539A-6\_COPY\_59\_86 (1-28) x AAT67618 (1-576)  
Qy 4 LeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAlaAlaValVallys 23  
Db 145 ATACAGCCCTACAGGACCAATTCAGCGCTTTAGATTCTCAAGAAAAACTCGTTAGCAA 204  
Qy 24 AlaAspAsn 26  
Db 205 TCGGATAAC 213  
RESULT 8  
AAT67891  
ID AAT67891 standard; DNA: 666 BP.  
XX  
AC AAT67891;  
XX  
XX 14-JUL-1997 (first entry)  
XX H. pylori secreted or periplasmic protein ORF 02cel0216orf1.  
XX  
XX Cytoplasmic; vaccine; prevention; treatment; infection; envelope;  
KW identification; binding compound; bacterium; life cycle; activator;  
KW bacteria; inhibitor; duodenal ulcer disease; chronic gastritis;  
KW diagnosis; ds.  
XX  
OS Helicobacter pylori.  
XX  
XX Key Location/Qualifiers  
FH 1..666  
FT /\*tag= a  
FT /note= "no stop codon given"  
XX  
XX W09640893-A1.  
XX 19-DEC-1996.  
XX 06-JUN-1996; 96WO-US09122.  
XX 01-APR-1996; 96US-0630405.  
XX 07-JUN-1995; 95US-0487032.  
XX (ASTR ) ASTRA AB.  
XX Berglindh OT, Smith D, Mellgaard BL;  
XX WPI; 1997-052306/05.  
XX P-PSDB; AAW20638.  
XX  
XX Helicobacter pylori nucleic acid sequences and related  
PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori  
PT infection, and to detect Helicobacter  
XX  
XX Claim 23; Page 770-771; 1481pp; English.  
XX  
XX The present sequence encodes a H. pylori secreted or periplasmic protein.  
CC The protein may be used in a vaccine to prevent or treat H. pylori  
CC infection or to identify H. pylori polypeptide binding compounds,  
CC useful as potential H. pylori life cycle activators or inhibitors.  
CC The genomic sequence of H. pylori (ATCC 55679) was determined from  
CC overlapping contigs generated by mechanically shearing the bacterial  
CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,



QY 21 ValVallysAlaAspAsnAlaA 28  
|||||  
Db 393 GTTGTTAAAGCGGATAACGCTGCT 416

RESULT 5  
AAD00561  
ID AAD00561 standard; DNA; 777 BP.

XX AC AAD00561;  
XX  
XX 29-AUG-2000 (first entry)

XX Streptococcus pyogenes strain KTL9 partial GRAB protein encoding DNA.  
XX GRAB protein; protein G related alpha2M binding protein; vaccine;  
KW alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;  
KW immune response; Streptococcus pyogenes infection; ds.  
OS Streptococcus pyogenes.

XX Key Location/Qualifiers  
FH 1..777  
FT CDS  
FT /\*tag= a  
FT /product= "GRAB protein"  
FT /note= "Does not include stop codon"  
FT /partial

XX WO200026240-A2.

XX 11-MAY-2000.

XX 02-NOV-1999; 99WO-GB03631.

XX 02-NOV-1998; 98GB-0023975.

XX (ACTI-) ACTINOVA LTD.

XX Bjorck LH, Rasmussen M;

XX WPI; 2000-365572/31.

XX P-PSDB; AAY71043.

XX New alpha2M binding protein for generating a protective immune response  
to group A streptococcus and purifying the binding protein -

XX Claim 13; Page 63-64; 67pp; English.

XX The patent discloses a new family of proteins termed GRAB (protein G  
CC related alpha2M binding protein) from Streptococcus pyogenes which have  
CC the ability to bind alpha2-macroglobulin (alpha2M) and show homology to  
CC protein G of group G Streptococcus. GRAB protein and peptides derived  
CC from it are used in vaccine compositions for generating a protective  
CC immune response against group A Streptococcus. Antibodies against GRAB  
CC are useful for treating Streptococcus pyogenes infections. The protein  
CC is also useful for purifying alpha2M from a sample. The present sequence  
CC is a DNA encoding partial GRAB protein from S. pyogenes strain KTL9.  
CC The protein has alpha2M binding region and is useful in vaccine  
CC composition.

XX SQ Sequence 777 BP; 269 A; 151 C; 178 G; 179 T; 0 other;

Alignment Scores:  
Pred. No.: 2,74e-13 Length: 777  
Score: 129.00 Matches: 28  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 21 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x AAD00561 (1-777)

QY 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluAlaA 20

Db 319 AGTCAGCGCCTTAGAAGCATTGGCGGATCAACAGACGCTTTACAATCAGAGAAGCTGCG 378  
|||||

QY 21 ValVallysAlaAspAsnAlaA 28  
|||||  
Db 379 GTTGTTAAAGCGGATAACGCTGCT 402

RESULT 6

AAD00563

ID AAD00563 standard; DNA; 853 BP.

XX AC AAD00563;

XX 29-AUG-2000 (first entry)

XX Streptococcus pyogenes strain AP49 partial GRAB protein encoding DNA.  
XX GRAB protein; protein G related alpha2M binding protein; vaccine;  
KW alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;  
KW immune response; Streptococcus pyogenes infection; ds.

XX Streptococcus pyogenes.

XX Key Location/Qualifiers  
FH 37..852  
FT CDS  
FT /\*tag= a  
FT /product= "GRAB protein"  
FT /partial

XX WO200026240-A2.

XX 11-MAY-2000.

XX 02-NOV-1999; 99WO-GB03631.

XX 02-NOV-1998; 98GB-0023975.

XX (ACTI-) ACTINOVA LTD.

XX Bjorck LH, Rasmussen M;

XX WPI; 2000-365572/31.

XX P-PSDB; AAY71045.

XX New alpha2M binding protein for generating a protective immune response  
to group A streptococcus and purifying the binding protein -

XX Claim 13; Page 64; 67pp; English.

XX The patent discloses a new family of proteins termed GRAB (protein G  
CC related alpha2M binding protein) from Streptococcus pyogenes which have  
CC the ability to bind alpha2-macroglobulin (alpha2M) and show homology to  
CC protein G of group G Streptococcus. GRAB protein and peptides derived  
CC from it are used in vaccine compositions for generating a protective  
CC immune response against group A Streptococcus. Antibodies against GRAB  
CC are useful for treating Streptococcus pyogenes infections. The protein  
CC is also useful for purifying alpha2M from a sample. The present sequence  
CC is a DNA encoding partial GRAB protein from S. pyogenes strain AP49.  
CC The protein has alpha2M binding region and is useful in vaccine  
CC composition.

XX SQ Sequence 853 BP; 295 A; 171 C; 197 G; 190 T; 0 other;

Alignment Scores:  
Pred. No.: 3.08e-13 Length: 853  
Score: 129.00 Matches: 28  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 21 Gaps: 0

US-09-847-539A-6\_COPY\_59\_86 (1-28) x AAD00563 (1-853)

```

FH Key      Location/Qualifiers
FT CDS      1..654
FT          /*tag= a
FT          /product= "GRAB protein"
FT sig_peptide 1..99
FT          /*tag= b
FT mat_peptide 100..651
FT          /*tag= c
FT          /product= "Mature GRAB protein"
XX
PN WO200026240-A2.
XX
XX 11-MAY-2000.
XX
PF 02-NOV-1999; 99WO-GB03631.
XX
PR 02-NOV-1998; 98GB-0023975.
XX
PA (ACTI-) ACTINOVA LTD.
XX
PI Bjorck LH, Rasmussen M;
XX
XX WPI; 2000-365572/31.
DR
DR P-PSDB; AAY71042.
XX
XX New alpha2M binding protein for generating a protective immune response
PT to group A streptococcus and purifying the binding protein
XX
XX Claim 13; Page 63; 67pp; English.
XX
XX The patent discloses a new family of proteins termed GRAB (protein G
CC related alpha2M binding protein) from Streptococcus pyogenes which have
CC the ability to bind alpha2-macroglobulin (alpha2M) and show homology to
CC protein G of group G Streptococcus. GRAB protein and peptides derived
CC from it are used in vaccine compositions for generating a protective
CC immune response against group A Streptococcus. Antibodies against GRAB
CC are useful for treating Streptococcus pyogenes infections. The protein
CC is also useful for purifying alpha2M from a sample. The present sequence
CC is a DNA encoding full-length GRAB protein from S. pyogenes strain SF370.
XX
XX Sequence 654 BP; 234 A; 118 C; 143 G; 159 T; 0 other;
SQ
Alignment Scores:
Pred. No.: 2 21e-13 Length: 654
Score: 129.00 Matches: 28
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 21 Gaps: 0
US-09-847-539A-6_COPY_59_86 (1-28) x AAD00560 (1-654)
Qy 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAla 20
Db 274 TCAGATGCGCTTAGAGCATTCAGCGATCAACAGACGCTTTACATCAGAGAGCTGCG 333
Qy 21 ValValLysAlaAspAsnAlaAla 28
Db 334 GTTGTAAAGCGGATACGCTGCT 357
RESULT 4
AAD00559
ID AAD00559 standard; DNA; 764 BP.
XX
AC AAD00559;
XX
DT 29-AUG-2000 (first entry)
XX
XX Streptococcus pyogenes strain SF370 GRAB protein encoding DNA.
XX
KW GRAB protein; protein G related alpha2M binding protein; vaccine;
KW alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;
KW immune response; Streptococcus pyogenes infection; ds.

```

```

XX Streptococcus pyogenes.
OS
XX
XX Location/Qualifiers
FH Key      60..713
FT CDS
FT          /*tag= a
FT          /product= "GRAB protein"
FT sig_peptide 60..158
FT          /*tag= b
FT mat_peptide 159..710
FT          /*tag= c
FT          /product= "Mature GRAB protein"
FT          complement (101..124)
FT          /*tag= d
FT primer_bind /bound_moiety= "Primer 1 represented in AAD00565"
FT          complement (101..127)
FT          /*tag= e
FT primer_bind /bound_moiety= "Primer 2 represented in AAD00566"
FT          complement (160..184)
FT          /*tag= f
FT primer_bind /bound_moiety= "Primer 3 represented in AAD00567"
FT          complement (563..594)
FT          /*tag= g
FT primer_bind /bound_moiety= "Primer 4 represented in AAD00568"
FT          complement (605..626)
FT          /*tag= h
FT          /bound_moiety= "Primer 5 represented in AAD00569"
XX
XX WO200026240-A2.
XX
XX 11-MAY-2000.
XX
XX 02-NOV-1999; 99WO-GB03631.
XX
XX 02-NOV-1998; 98GB-0023975.
XX
XX (ACTI-) ACTINOVA LTD.
XX
XX Bjorck LH, Rasmussen M;
XX
XX WPI; 2000-365572/31.
DR
DR P-PSDB; AAY71042.
XX
XX New alpha2M binding protein for generating a protective immune response
PT to group A streptococcus and purifying the binding protein
XX
XX Example 1; Fig 2B; 67pp; English.
XX
XX The patent discloses a new family of proteins termed GRAB (protein G
CC related alpha2M binding protein) from Streptococcus pyogenes which have
CC the ability to bind alpha2-macroglobulin (alpha2M) and show homology to
CC protein G of group G Streptococcus. GRAB protein and peptides derived
CC from it are used in vaccine compositions for generating a protective
CC immune response against group A Streptococcus. Antibodies against GRAB
CC are useful for treating Streptococcus pyogenes infections. The protein
CC is also useful for purifying alpha2M from a sample. The present sequence
CC is a DNA encoding full-length GRAB protein from S. pyogenes strain SF370.
XX
XX Sequence 764 BP; 279 A; 131 C; 159 G; 195 T; 0 other;
SQ
Alignment Scores:
Pred. No.: 2 68e-13 Length: 764
Score: 129.00 Matches: 28
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 21 Gaps: 0
US-09-847-539A-6_COPY_59_86 (1-28) x AAD00559 (1-764)
Qy 1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAla 20
Db 333 TCAGATGCGCTTAGAGCATTCAGCGATCAACAGACGCTTTACATCAGAGAGCTGCG 392

```

```

FT      /*tag= a
FT      /product= "GRAB protein"
FT      /partial
XX
PN      WO200026240-A2.
XX
PD      11-MAY-2000.
XX
PF      02-NOV-1999; 99WO-GB03631.
XX
PR      02-NOV-1998; 98GB-0023975.
XX
PA      (ACTI-) ACTINOVA LTD.
XX
PI      Bjorck LH, Rasmussen M;
XX
XX      WPI; 2000-365572/31.
DR      P-PSDB; AAY71044.
DR
XX
PT      New alpha2M binding protein for generating a protective immune response
PT      to group A streptococcus and purifying the binding protein
XX
PS      Claim 13; Page 64; 67pp; English.
XX
CC      The patent discloses a new family of proteins termed GRAB (protein G
CC      related alpha2M binding protein) from Streptococcus pyogenes which have
CC      the ability to bind alpha2-macroglobulin (alpha2M) and show homology to
CC      protein G of group G Streptococcus. GRAB protein and peptides derived
CC      from it are used in vaccine compositions for generating a protective
CC      immune response against group A Streptococcus. Antibodies against GRAB
CC      are useful for treating Streptococcus pyogenes infections. The protein
CC      is also useful for purifying alpha2M from a sample. The present sequence
CC      is a DNA encoding partial GRAB protein from S. pyogenes strain AP1.
CC      The protein has alpha2M binding region and is useful in vaccine
CC      composition.
XX
SQ      Sequence 469 BP; 180 A; 90 C; 99 G; 100 T; 0 other;

Alignment Scores:
Pred. No.: 1.45e-13 Length: 469
Score: 129.00 Matches: 28
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 21 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x AAD00562 (1-469)
QY      1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAla 20
DB      172 TCAGATGCGCTTAGAGCATTTAGCGGATCAACAGACGCTTTACATCAGAGAAGCTGCG 231
QY      21 ValValLysAlaAspAsnAlaAla 28
DB      232 GTTGTAAAGCGGATACGCTGCT 255

RESULT 2
AAD00564
ID      AAD00564 standard; DNA; 504 BP.
XX
AC      AAD00564;
XX
DT      29-AUG-2000 (first entry)
XX
DE      Streptococcus pyogenes strain KTL3 partial GRAB protein encoding DNA.
XX/
KW      GRAB protein; protein G related alpha2M binding protein; vaccine;
KW      alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;
KW      immune response; Streptococcus pyogenes infection; ds.
XX
OS      Streptococcus pyogenes.
XX
FH      Key Location/Qualifiers

FT      CDS 1..504
FT      /*tag= a
FT      /product= "GRAB protein"
FT      /partial
XX
PN      WO200026240-A2.
XX
PD      11-MAY-2000.
XX
PF      02-NOV-1999; 99WO-GB03631.
XX
PR      02-NOV-1998; 98GB-0023975.
XX
PA      (ACTI-) ACTINOVA LTD.
XX
PI      Bjorck LH, Rasmussen M;
XX
XX      WPI; 2000-365572/31.
DR      P-PSDB; AAY71046.
DR
XX
PT      New alpha2M binding protein for generating a protective immune response
PT      to group A streptococcus and purifying the binding protein
XX
PS      Claim 13; Page 65; 67pp; English.
XX
CC      The patent discloses a new family of proteins termed GRAB (protein G
CC      related alpha2M binding protein) from Streptococcus pyogenes which have
CC      the ability to bind alpha2-macroglobulin (alpha2M) and show homology to
CC      protein G of group G Streptococcus. GRAB protein and peptides derived
CC      from it are used in vaccine compositions for generating a protective
CC      immune response against group A Streptococcus. Antibodies against GRAB
CC      are useful for treating Streptococcus pyogenes infections. The protein
CC      is also useful for purifying alpha2M from a sample. The present sequence
CC      is a DNA encoding partial GRAB protein from S. pyogenes strain KTL3.
CC      The protein has alpha2M binding region and is useful in vaccine
CC      composition.
XX
SQ      Sequence 504 BP; 188 A; 97 C; 108 G; 111 T; 0 other;

Alignment Scores:
Pred. No.: 1.59e-13 Length: 504
Score: 129.00 Matches: 28
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 21 Gaps: 0

US-09-847-539A-6_COPY_59_86 (1-28) x AAD00564 (1-504)
QY      1 SerAspAlaLeuGluAlaLeuAlaAspGlnThrAspAlaLeuGlnSerGluGluAla 20
DB      208 TCAGATGCGCTTAGAGCATTTAGCGGATCAACAGACGCTTTACATCAGAGAAGCTGCG 267
QY      21 ValValLysAlaAspAsnAlaAla 28
DB      268 GTTGTAAAGCGGATACGCTGCT 291

RESULT 3
AAD00560
ID      AAD00560 standard; DNA; 654 BP.
XX
AC      AAD00560;
XX
DT      29-AUG-2000 (first entry)
XX
DE      Streptococcus pyogenes strain SF370 GRAB protein coding region.
XX/
KW      GRAB protein; protein G related alpha2M binding protein; vaccine;
KW      alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;
KW      immune response; Streptococcus pyogenes infection; ds.
XX
OS      Streptococcus pyogenes.
XX
FH      Key Location/Qualifiers

```

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 13, 2002, 02:10:56 : Search time 40.4278 Seconds  
(without alignments)  
1189.122 Million cell updates/sec

Title: US-09-847-539a-6\_COPY\_59\_86

Perfect score: 129  
Sequence: 1 SDALALADQDTALQSEAAVVKADNAA 28

Scoring table: BLOSUM62  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters: -DEV=xlh

-MODEL=frame\_p2n.model -DEV=xlh  
-Q/cgn2\_1/USPTO\_spool/US09847539/runat\_10102002\_092548\_3413/app\_query.fasta\_1.526  
-DB=N\_Geneseq\_032802 -QFMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blom62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=40  
-MODE=LOCAL -OUTFMT=ptc -NORW=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09847539\_ECGN\_1.1.48@runat\_10102002\_092548\_3413 -NCPU=6 -ICPU=3  
-NO\_XLPHY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

N\_Geneseq\_032802.\*  
1: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	129	100.0	469	21	AAD00562	Streptococcus pyog
2	129	100.0	504	21	AAD00564	Streptococcus pyog
3	129	100.0	654	21	AAD00560	Streptococcus pyog
4	129	100.0	764	21	AAD00559	Streptococcus pyog
5	129	100.0	777	21	AAD00561	Streptococcus pyog
6	129	100.0	853	21	AAD00563	Streptococcus pyog
7	57	44.2	576	18	AAT67618	H. pylori secreted
8	57	44.2	666	18	AAT67891	H. pylori secreted
9	57	44.2	1239	19	AA30461	H. pylori secreted
C 10	54	41.9	3349	23	ABL13527	Drosophila melanog
C 11	54	41.9	6398	23	ABL13526	Drosophila melanog
C 12	52	40.3	1002	21	AAC54228	Arabidopsis thalia
C 13	52	40.3	1230	21	AAC33266	Arabidopsis thalia
C 14	52	40.3	16135	23	ABL25211	Drosophila melanog
C 15	52	40.3	18603	23	ABL25210	Drosophila melanog
C 16	51	39.5	1960	23	ABL12683	Drosophila melanog
C 17	51	39.5	2618	23	ABL27360	Drosophila melanog
C 18	51	39.5	4221	23	ABL12682	Drosophila melanog
19	51	39.5	2944528	24	ABA03041	Listeria monocytog
20	50	38.8	3711	23	AA594455	DNA encoding novel
C 21	48	37.2	1616	23	ABL25849	Drosophila melanog
C 22	48	37.2	4229	23	ABL25848	Drosophila melanog
C 23	47	36.4	4403765	22	AA199683	Mycobacterium tube
C 24	47	36.4	4411529	22	AA199682	Mycobacterium tube
25	46	35.7	636	21	AAF11343	Aspergillus niger
26	46	35.7	2115	22	AAH65520	C glutamicum codin
C 27	46	35.7	2160	22	AAH65521	C glutamicum codin
C 28	46	35.7	69936	21	AAH81479	N. meningitidis pa
C 29	46	35.7	349980	21	AAF21607	Neisseria meningit
30	46	35.7	349980	22	AAH68525	C glutamicum codin
31	46	35.7	1437668	21	AAH81490	N. meningitidis B
32	45	34.9	417	21	AAZ56335	Escherichia coli f
C 33	45	34.9	461	21	AAF08476	Fusarium venenatum
C 34	45	34.9	1365	21	AAZ56356	Escherichia coli f
C 35	45	34.9	1368	21	AAZ56355	Escherichia coli f
C 36	45	34.9	1380	21	AAZ56351	Escherichia coli f
37	45	34.9	1380	21	AAZ56376	Escherichia coli f
38	45	34.9	1383	21	AAZ56340	Escherichia coli f
39	45	34.9	1383	21	AAZ56368	Escherichia coli f
C 40	45	34.9	1773	21	AAZ56162	Coding sequence of
C 41	45	34.9	2145	21	AAZ56364	Escherichia coli f
C 42	45	34.9	2599	21	AAZ56382	Escherichia coli f
C 43	45	34.9	4621	20	AAV99919	Murine pcip gene (
C 44	45	34.9	4860	20	AAZ26000	Murine p/cip gene,
C 45	45	34.9	5091	23	ABL07051	Drosophila melanog

ALIGNMENTS

RESULT 1  
AAD00562  
ID AAD00562 standard; DNA: 469 BP.  
XX  
AC AAD00562;  
XX  
DT 29-AUG-2000 (first entry)  
XX  
DE Streptococcus pyogenes strain APl partial GRAB protein encoding DNA.  
XX  
KW GRAB protein; protein G related alpha2M binding protein; vaccine;  
KW alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;  
KW Immune response; Streptococcus pyogenes infection; ds.  
XX  
OS Streptococcus pyogenes.  
XX  
FH Key Location/Qualifiers  
FT CDS 1..468

THIS PAGE BLANK (uspto)



```
QY   26 EKALRNEERAIDELKKQAIEDKEATTAIAEASDSDALEALADOTDALQSEEAHVVKADNA 85
      | : | : | : ||| : | : | : | : | : | : | : | : | : | : | : | : | :
Db    27 ENNVAKENNTESGEKQNQTVAETTTTSVEAKETFPPVPTKETTPAVOPEVAAEVSSSA 86
      | : | : | : ||| : | : | : | : | : | : | : | : | : | : | : | : | :
QY   86 ASDALEALADOTDALQSEEEVVQSNDNAAASDAWEKAATPIALDVK 130
      | : | : | : ||| : | : | : | : | : | : | : | : | : | : | : | : | :
Db    87 DAGEAIVPAPEKVENAENAEAKVEAVAAPAEKVEVAEAEKK 131
      | : | : | : ||| : | : | : | : | : | : | : | : | : | : | : | : | :

RESULT 34
C90769
C:Species: Escherichia coli
C:title: probable tail fiber protein [imported] - Escherichia coli (strain O157:H7, substrain RIM)
C:date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
R:Accession: C90769
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A>Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Cross-references: GB:BA000007; PIDN:BA34546.1; PID:g13360583; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RMD 0509952
C:Genetics:
A:Gene: ECs1123

Query Match          9.4%; Score 73; DB 2; Length 439;
Best Local Similarity 35.3%; Pred. No. 4.1;
Matches 24; Conservative 7; Mismatches 37; Indels 0; Gaps 0;

QY   50 ATTAIEAESDALEALADOTDALQSEEAHVVKADNAASDALEALADOTDALQSEEAHVQ 109
      | : | : | : ||| : | : | : | : | : | : | : | : | : | : | : | : | :
Db    150 ADTSGADSARSARQAESAAMAAKOSEEASSSSAAAOKASESSQSAEAEELSRRKTASA 209
      | : | : | : ||| : | : | : | : | : | : | : | : | : | : | : | : | :

QY   110 SDNAASDA 117
       | : | : | 
Db    210 AGNAARDA 217

RESULT 35
D84325
Htri7 transducer [imported] - Halobacterium sp. NRC-1
C:Species: Halobacterium sp. NRC-1
C:date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
C:Accession: D84325
R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.;
Leithausner, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jabid,
Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li
A>Title: Genome sequence of Halobacterium species NRC-1.
A:Reference number: AB4160; MUID:20504483
A:Accession: D84325
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-536 <STO>
A:Cross-references: GB:AB004437; MID:g10581193; PIDN:AAG19968.1; GSPDB:GN00138
C:Genetics:
A:Gene: htri7
C:Superfamily: Halobacterium salinarum transducer protein htri

Query Match          9.4%; Score 73; DB 2; Length 536;
Best Local Similarity 27.0%; Pred. No. 5;
Matches 24; Conservative 12; Mismatches 53; Indels 0; Gaps 0;

QY   39 ELKKQAIEDKCATTAIEAESSDALEALADOTDALQSEEAHVVKADNAASDALEALADOTD 98
      | : | : | : ||| : | : | : | : | : | : | : | : | : | : | : | : | :
Db    368 EVKLAEETNEATEIETFLVDNLNSTSQSVYDDSEQMAATVSSGTCTVENALTAALEEIGD 427
```

A:Gene: SGD:SRP40  
A:Cross-references: SGD:S0001800; MIPS:YKR092c  
A:Map position: 11R

Query Match 9.7%; Score 75; DB 2; Length 406;  
Best Local Similarity 18.5%; Pred. No. 2.3;  
Matches 15; Conservative 29; Mismatches 37; Indels 0; Gaps 0;

QY 37 IDELKKQAIEDKEATTATAEAASSDALEALADOTDALQSEAAVVKADNAASDALEALADQ 96  
DB 8 VDEVPKLSVKEKEIEKSS 67

QY 97 TDALQSEAAEVQSDNAASDA 117

DB 68 SDSDSSESSSSSSSSSSSS 88

RESULT 30

AD2395  
two-component sensor histidine kinase alr4716 [Imported] - Anabaena sp. (strain PCC 7120)  
C:Species: Anabaena sp.  
A:Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120  
C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 11-Jan-2002  
C:Accession: AD2395  
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kurlitz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, R.; Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S. DNA Res. 8, 205-213, 2001  
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena sp. strain PCC 7120  
A:Reference number: AB1807; MUID:21595285; PMID:11759840  
A:Accession: AD2395  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-468 <KUR>  
A:Cross-references: GB:BA000019; PIDN:BA076415.1; PID:g17133853; GSPDB:GN00179  
A:Experimental source: strain PCC 7120  
C:Genetics:  
A:Gene: alr4716

Query Match 9.7%; Score 75; DB 2; Length 468;  
Best Local Similarity 25.6%; Pred. No. 2.6;  
Matches 21; Conservative 19; Mismatches 42; Indels 0; Gaps 0;

QY 34 ERAIDELKKQATIEDKEATTATAEAASSDALEALADOTDALQSEAAVVKADNAASDALEAL 93  
DB 129 EAANNQLXIEIERTAEAAQLQIKADLENRVVEETNLOQSNLASLATAAKAQAEKL 188

QY 94 ADQTDALQSEAAEVQSDNAAS 115

DB 189 OVALNDLANAQVLIQTCKMSS 210

RESULT 31

G96552  
unknown protein, 66348-64527 [Imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C:Accession: G96552  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, J.; Chiu, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Creasy, T.H.; Dewar, K.; Hansen, N.F.; Hughes, B.; Huizar, L. Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.; C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: AB6141; MUID:21016719  
A:Accession: G96552  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-509 <STO>  
A:Cross-references: GB:AE005173; NID:g10092363; PIDN:AAG12772.1; GSPDB:GN00141

C:Genetics:  
A:Gene: F5D21.18  
A:Map position: 1

Query Match 9.7%; Score 75; DB 2; Length 509;  
Best Local Similarity 26.4%; Pred. No. 2.9;  
Matches 24; Conservative 18; Mismatches 49; Indels 0; Gaps 0;

QY 46 EKEATTATAEAASSDALEALADOTDALQSEAAVVKADNAASDALEALADQTDALQSEEA 105  
DB 131 ERAPTTSSDNPSPKLDSSVANQNGSAMDEGDEQDDPPHKKLKQLDCLTSVAVKEEB 190

QY 106 EYVQSDNRAASDAWEKAATPIALDVKKTKDK 136

DB 191 EPEQVLPSEAMVVEAATLVASAANKSKSK 221

RESULT 32

G85631  
hypothetical protein 21382 [Imported] - Escherichia coli (strain O157:H7, substrain E  
C:Species: Escherichia coli  
C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001  
C:Accession: G85631  
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May  
Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apoda  
Nature 409, 529-533, 2001  
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
A:Reference number: A85480; MUID:21074935; PMID:11206551  
A:Accession: G85631  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-375 <STO>  
A:Cross-references: GB:AE005174; NID:g12514226; PIDN:AAG55515.1; GSPDB:GN00145; UWCP:  
A:Experimental source: strain O157:H7, substrain EDL933  
C:Genetics:  
A:Gene: Z1382

Query Match 9.5%; Score 74; DB 2; Length 375;  
Best Local Similarity 35.3%; Pred. No. 2.7;  
Matches 24; Conservative 8; Mismatches 36; Indels 0; Gaps 0;

QY 50 ATTAETAASSDALEALADOTDALQSEAAVVKADNAASDALEALADQTDALQSEAAEVVQ 109  
DB 150 ADTSAGDASARQAESAANAQAQSEEAASSSSASANAQAQSESSQSSAANAELSKKTAESA 209

QY 110 SDNAASDA 117

DB 210 AGNAARDA 217

RESULT 33

T46231  
hypothetical protein T9C5.130 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 04-Feb-2000  
C:Accession: T46231  
R:Rieger, M.; Gabel, C.; Mueller-Auer, S.; Schaefer, M.; Zipp, M.; Mewes, H.W.; Lemck  
submitted to the Protein Sequence database, December 1999  
A:Reference number: Z23026  
A:Accession: T46231  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-166 <RIE>  
A:Cross-references: EMBL:AL132964  
A:Experimental source: cultivar Columbia; BAC clone T9C5  
C:Genetics:  
A:Map position: 3  
A:Introns: 32/3  
A:Note: T9C5.130  
Query Match 9.4%; Score 73; DB 2; Length 166;  
Best Local Similarity 21.9%; Pred. No. 1.4;  
Matches 23; Conservative 22; Mismatches 60; Indels 0; Gaps 0;







A:Molecule type: DNA

A:Residues: 1-507 <FUE>

A:Cross-references: GB:X16421; EMBL:M26048; NID:g43333; PIDN:CAA34442.1; PID:g43334

A:Note: the authors translated the codon CGT for residues 221 and 223 as Lys

Query Match 10.2%; Score 79; DB 2; Length 507;

Best Local Similarity 16.9%; Pred. No. 1;

Matches 21; Conservative 30; Mismatches 73; Indels 0; Gaps 0;

Qy 26 EKALRNEERAIDELKKAOTEDKATTAIAAASDDALEALADQTDALQSEAAVVKADNA 85

Db 218 EQAIRQEARLAERQAQAEKAEKAREQAAQAQATQALSSASTTTSSSSAAQSSE 277

Qy 86 ASDALEALADQTDALQSEAAVWQSDNAASDAWEKAATPIALDVYKTKDKPKVVKKEEQ 145

Db 278 ESKAPESSTTEESTSTTSSTTSSSTSSSTSSSTSSSTSSSTSSSTSSSTSSSTSSSTSS 337

Qy 146 NVNT 149

Db 338 SSNT 341

RESULT 19

S66805

hypothetical protein YOL109w - yeast (*Saccharomyces cerevisiae*)

N:Alternate names: hypothetical protein O0738

C:Species: *Saccharomyces cerevisiae*

C>Date: 12-Jul-1996 #sequence\_revision 12-Jul-1996 #text\_change 06-Feb-1998

C:Accession: S66805

R:Durant, P.; Hilger, F.; Portetelle, D.; Vandenbol, M.

submitted to the Protein Sequence Database, July 1996

A:Reference number: S66791

A:Molecule type: DNA

A:Residues: 1-113 <DUR>

A:Cross-references: EMBL:Z74851; NID:g1419974; PID:e251901; PID:g1419975; MIPS:YOL109w

A:Experimental source: strain S288C

C:Genetics:

A:Gene: SGD:ZEO1

A:Cross-references: SGD:S0005469; MIPS:YOL109w

A:Map position: 15L

Query Match

Best Local Similarity 10.0%; Score 78; DB 2; Length 113;

Matches 19; Conservative 10; Mismatches 22; Indels 0; Gaps 0;

Qy 81 KADNAASDALEALADQTDALQSEAAVWQSDNAASDAWEKAATPIALDVYK 131

Db 7 KAETAAQDVQOKLEETKESLQKQGVKEQAESIDNLKNEATPEAEQVK 57

RESULT 20

G90907

probable tail fiber protein [Imported] - *Escherichia coli* (strain O157:H7, substrain RIM

C:Species: *Escherichia coli*

C>Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 18-Jul-2001

C:Accession: G90907

R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.

gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001

A:Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and gen

A:Reference number: A99629; MUID:21156231; PMID:11258796

A:Accession: G90907

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-407 <HAY>

A:Cross-references: GB:BA000007; PIDN:BA335654.1; PID:g13361697; GSPDB:GN00154

A:Experimental source: strain O157:H7, substrain RIMD 0509952

C:Genetics:

A:Gene: ECs2231

Query Match

Best Local Similarity 10.0%; Score 78; DB 2; Length 407;

Matches 19; Conservative 10; Mismatches 22; Indels 0; Gaps 0;

Matches 25; Conservative 7; Mismatches 36; Indels 0; Gaps 0;

Qy 50 ATTAIEAASDALEALADQTDALQSEAAVVKADNAASDALEALADQTDALQSEAAEVVQ 109

Db 148 ADTSAGDASEARQAESAASAAKQSEASSSASAAQAQKASESSQSAADAELSKKTAE 207

Qy 110 SDNAASDA 117

Db 208 AGNAARDA 215

RESULT 21

E90968

probable tail fiber protein [Imported] - *Escherichia coli* (strain O157:H7, substrain

C:Species: *Escherichia coli*

C>Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 18-Jul-2001

C:Accession: E90968

R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C

gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001

A:Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and g

A:Reference number: A99629; MUID:21156231; PMID:11258796

A:Accession: E90968

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-437 <HAY>

A:Cross-references: PIDN:BA336140.1; PID:g13362185; GSPDB:GN00154

A:Experimental source: strain O157:H7, substrain RIMD 0509952

C:Genetics:

A:Gene: Ecs2717

Query Match

Best Local Similarity 10.0%; Score 78; DB 2; Length 437;

Matches 25; Conservative 7; Mismatches 36; Indels 0; Gaps 0;

Qy 50 ATTAIEAASDALEALADQTDALQSEAAVVKADNAASDALEALADQTDALQSEAAEVVQ 109

Db 148 ADTSAGDASEARQAESAASAAKQSEASSSASAAQAQKASESSQSAADAELSKKTAE 207

Qy 110 SDNAASDA 117

Db 208 AGNAARDA 215

RESULT 22

E85816

probable tail fiber protein of prophage CP-933U Z3074 [Imported] - *Escherichia coli* (

C:Species: *Escherichia coli*

C>Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001

C:Accession: E85816

R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May

iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamouis, K.; Apoda

Nature 409, 529-533, 2001

A:Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.

A:Reference number: A85480; MUID:21074935; PMID:11206551

A:Accession: E85816

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-439 <STO>

A:Cross-references: GB:AE005174; NID:gl2516091; PIDN:AAG56993.1; GSPDB:GN00145; UWGP:

A:Experimental source: strain O157:H7, substrain EDL933

C:Genetics:

A:Gene: Z3074

Query Match

Best Local Similarity 10.0%; Score 78; DB 2; Length 439;

Matches 25; Conservative 7; Mismatches 36; Indels 0; Gaps 0;

Qy 50 ATTAIEAASDALEALADQTDALQSEAAVVKADNAASDALEALADQTDALQSEAAEVVQ 109

Db 150 ADTSAGDASEARQAESAASAAKQSEASSSASAAQAQKASESSQSAADAELSKKTAE 209

Qy 110 SDNAASDA 117







A:Accession: S00128  
A:Molecule type: DNA  
A:Residues: 1-593 <OLS>  
A:Cross-references: EMBL:X06173; NID:g47084; PIDN:CAA29540.1; PID:g47085  
A>Note: the source is designated as Streptococcus G148  
R:Sjoerling, U.; Falkenberg, C.; Nielsen, E.; Akerstrom, B.; Bjoerck, L.  
J. Immunol. 140, 1595-1599, 1988  
A:Title: Isolation and characterization of a 14-kDa albumin-binding fragment of streptococcal protein G.  
A:Reference number: A27604; MUID:88154455  
A:Accession: A27604  
A:Molecule type: protein  
A:Residues: 62-101 <SJO>  
R:Guss, B.; Eliasson, M.; Olsson, A.; Uhlen, M.; Frej, A.K.; Jornvall, H.; Flock, J.I.; EMBO J. 5, 1567-1575, 1986  
A:Title: Structure of the IgG-binding regions of streptococcal protein G.  
A:Reference number: A26314; MUID:86300657  
A:Accession: A26314  
A:Molecule type: DNA  
A:Residues: 114-593 <GUS>  
A:Cross-references: GB:X04015; NID:g47071; PIDN:CAA27638.1; PID:g47072  
C:Function:  
A:Description: it is part of the cell wall structure of group G streptococci and is involved in the adsorption of antibodies to the cell wall.  
C:Superfamily: M5 protein  
F:1-33/Domain: signal sequence #status predicted <SIG>  
F:34-593/Product: protein G #status experimental <MAT>  
F:34-116/Domain: alanine-rich <ALA>  
F:117-290/Domain: AB duplication <DUP1>  
F:117-140/Region: A repeat  
F:141-191/Region: B  
F:192-215/Region: A repeat  
F:216-266/Region: B  
F:267-290/Region: A repeat  
F:303-497/Domain: IgG binding <IGB>  
F:303-357/Region: C repeat  
F:358-372/Region: D  
F:373-427/Region: C repeat  
F:428-442/Region: D  
F:443-497/Region: C repeat  
F:498-567/Domain: proline-rich <PRO>  
F:568-593/Domain: carboxyl-terminal <CTD>

Query Match 19.9%; Score 155; DB 2; Length 593;  
Best Local Similarity 56.1%; Pred. No. 5.8e-09;  
Matches 37; Conservative 4; Mismatches 25; Indels 0; Gaps 0;

QY 1 VDSPIEQPIIPNGGTLNLGNAPKALRNEERAEIDELKKAIEDKEATTAIEAASD 60  
DB 34 VDSPIEDTPIIRNGGELTLLGNSETTLALRNEESATDLTAATAAVDTVAANAAGAA 93

QY 61 ALEA 66  
DB 94 AWEAAA 99

RESULT 3  
A44801  
albumin-binding protein precursor - Streptococcus sp. (fragment)  
C:Species: Streptococcus sp.  
C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999  
C:Accession: A44801  
R:Sjoerling, U.  
Infect. Immun. 60, 3601-3608, 1992  
A:Title: Isolation and molecular characterization of a novel albumin-binding protein from Streptococcus sp.  
A:Reference number: A44801; MUID:92363555  
A:Molecule type: DNA; protein  
A:Residues: 1-323 <SJO>  
A:Cross-references: GB:M95520; NID:g153554; PIDN:AAA26847.1; PID:g153555  
A:Experimental source: group G, strain DG12  
A>Note: sequence extracted from NCBI backbone (NCBIN:110938, NCBI:110939)  
F:12-323/Product: albumin-binding protein (fragment) #status experimental <MAT>

Query Match 16.2%; Score 126; DB 2; Length 323;  
Best Local Similarity 42.6%; Pred. No. 4.5e-06;  
Matches 26; Conservative 14; Mismatches 21; Indels 0; Gaps 0;

QY 62 LEALADQTDALQSEAAVVKADNAASDALEALADQTDALQSEAAEVQSDNAASDAWEKA 121  
DB 62 LSAQADQIVSAQADNEATTKAEEDSSKAWAAADQANTAKADELAKAEKSSDAWEKA 121

QY 122 A 122  
DB 122 A 122

RESULT 4  
OXBP11  
hypothetical protein 401 - phage lambda  
N:Alternate names: orf-401; orf401  
C:Species: phage lambda  
C:Date: 13-Jun-1983 #sequence\_revision 13-Jun-1983 #text\_change 23-Jul-1999  
C:Accession: G43010; D43016; A04389  
R:Daniels, D.  
submitted to the Nucleic Acid Sequence Database, September 1982  
A:Reference number: A94614  
A:Accession: G43010  
A:Molecule type: DNA  
A:Residues: 1-401 <DAN>  
R:Sanger, F.; Coulson, A.R.; Hong, G.F.; Hill, D.F.; Petersen, G.B.  
J. Mol. Biol. 162, 729-773, 1982  
A:Title: Nucleotide sequence of bacteriophage lambda DNA.  
A:Reference number: A92891; MUID:83189071  
A:Accession: D43016  
A:Molecule type: DNA  
A:Residues: 1-401 <SAN>  
A:Cross-references: GB:J02459; GB:M17233; GB:M24325; GB:V00636; GB:X00906; NID:g21510  
C:Genetics:  
A:Map position: 40.51-42.99  
C:Superfamily: phage lambda hypothetical protein 401

Query Match 12.7%; Score 99; DB 1; Length 401;  
Best Local Similarity 32.9%; Pred. No. 0.0052;  
Matches 26; Conservative 16; Mismatches 37; Indels 0; Gaps 0;

QY 50 ATTATBAASSDALEALADQTDALQSEAAVVKADNAASDALEALADQTDALQSEAAEVQ 109  
DB 190 AISSSKNAATSAAGAAKTSETNAAASQQAATSAATKASEAATSAARDAAVASKAASS 249

QY 110 SDNAASDAWEKAATPTALD 128  
DB 250 ETNASSAGRAASSATAAE 268

RESULT 5  
D90734  
probable tail fiber protein [imported] - Escherichia coli (strain O157:H7, substrain C:Species: Escherichia coli  
C:Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 18-Jul-2001  
C:Accession: D90734  
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.  
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
DNA Res. 8, 11-22, 2001  
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and g  
A:Reference number: A99629; MUID:21156231; PMID:11258796  
A:Accession: D90734  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-438 <HAY>  
A:Cross-references: GB:BA000007; PIDN:BA34267.1; PID:g13360303; GSPDB:GN00154  
A:Experimental source: strain O157:H7, substrain RMD 0509952  
C:Genetics:  
A:Gene: ECs0844

Query Match 12.4%; Score 96; DB 2; Length 438;

Result No.	Score	Query		Length	DB	ID	Description
		Match					
1	155	19.9	448	2	A24496		IgG-binding protein G precursor
2	155	19.9	593	2	S00128		albumin-binding protein
3	126	16.2	323	2	A4801		hypothetical protein
4	99	12.7	401	1	QXBPL		probable tail fiber
5	96	12.4	438	2	S09734		probable tail comp
6	96	12.4	440	2	F85584		plasma protein receptor
7	91	11.7	413	2	S55890		probable tail fiber
8	89	11.5	437	2	H90854		probable tail fiber
9	88	11.3	407	1	E0BEQ3		immediate-early protein
10	88	11.3	971	2	S90835		probable tail fiber
11	88	11.3	973	2	C85693		probable membrane
12	85	10.9	664	2	S42574		streptococcal surface
13	85	10.9	1122	2	G64887		probable tail fiber
14	80	10.3	269	2	S48444		probable temperature
15	80	10.3	347	2	E83525		ToIA protein PA097
16	80	10.3	1034	2	AB0551		exonuclease SbcC
17	79	10.2	389	2	S53975		probable membrane
18	79	10.2	507	2	S05542		hypothetical protein
19	78	10.0	113	2	S66805		hypothetical protein
20	78	10.0	407	2	G50907		probable tail fiber
21	78	10.0	437	2	E90968		probable tail fiber
22	78	10.0	439	2	E85816		probable tail fiber
23	78	10.0	771	1	A33430		h-caldesmon - chick
24	77	9.9	1110	2	I51116		NF-180 - sea lamp
25	77	9.9	1133	2	T22976		hypothetical protein
26	76	9.8	490	2	I41024		colicin 10 - Escherichia
27	76	9.8	725	2	A47168		cardiac morphogen
28	76	9.8	1290	2	A55094		chromosomal protein
29	75	9.7	406	2	S38170		SRP40 protein - yeast

Ellenstrom, A.; Eliasson, M.; Guss, B.; Nilsson, B.; Hellman, U.; Lindberg, M.; Uhlen, P.; Olsson, J. *Eur. J. Biochem.* 168, 319-324, 1987



**THIS PAGE BLANK (USPTO)**

```
CC -|- ALTERNATIVE PRODUCTS: THE PRODUCTION OF THE TWO PROTEIN PRODUCTS
CC FROM THIS REGION IS DUE TO PROGRAMED RIBOSOMAL FRAMESHIFTING.
CC -|- SIMILARITY: PARTIAL TO PHAGE T4 DNA POLYMERASE ACCESSORY PROTEIN
CC 44 AND TO HUMAN ACTIVATOR 1, 37 AND 40 KDA SUBUNITS.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X04487; CAA28174.1; -.
DR EMBL: X04487; CAA28175.1; ALT_INIT.
DR EMBL: X04275; CAA27827.1; -.
DR EMBL: AE000153; AAC73572.1; -.
DR EMBL: U82664; ABA0224.1; -.
DR EMBL: M38777; AAA23457.1; -.
DR PIR: A25549; DJEC3G.
DR ECODDBASE: H052.0; 6TH EDITION.
DR ECODDBASE: H080.0; 6TH EDITION.
DR Ecogene: EG10245; dnax.
DR InterPro: IPR003959; AAA_subfam.
DR InterPro: IPR000862; RFC.
DR Pfam: PF00004; AAA; 1.
DR Transferase: DNA-directed DNA polymerase; DNA replication;
DR ATP-binding; Ribosomal frameshift; Complete proteome.
DR CHAIN 1 643 DNA POLYMERASE III SUBUNIT TAU.
FT CHAIN 1 431 DNA POLYMERASE III SUBUNIT GAMMA.
FT NP_BIND 45 52 ATP (POTENTIAL).
SQ SEQUENCE 643 AA; 71137 MW; D2028BD99E375150 CRC64;

Query Match      8.8%; Score 68; DB 1; Length 643;
Best Local Similarity 36.5%; Pred. No. 7.3;
Matches 19; Conservative 9; Mismatches 24; Indels 0; Gaps 0;

QY 51 TTAIEAASDDALEALADOTDALQSEAAVVKADNAASDALEALADOTDALQS 102
DB 408 TTSQLARQQLQVQGATKAKKSEPAATRAPVNNALERLASVTDKRVQA 459

RESULT 40
ID DNAK_CHLMU STANDARD; PRT; 654 AA.
AC P56836.
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Chaperone protein dnaK (Heat shock protein 70) (Heat shock 70 kDa
DE protein) (HSP70) (75 kDa membrane protein).
GN DNAK OR TC0675.
OS Chlamydia muridarum.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=833560;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MOPN / NIGG;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K.,
RA Bass S., Linher K., Weidman J., Khouri H., Craven B., Bowman C.,
RA Dodson R., Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G.,
RA Salzberg S.L., Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RT pneumoniae AR39."
RL Nucleic Acids Res. 28:1397-1406(2000).
RN [2]
RP SEQUENCE OF 1-241 FROM N.A.
RC STRAIN=MOPN;
RX MEDLINE=91072247; PubMed=2254267;
RA Engel J.N., Pollack J., Perara E., Ganem D.;
RT "Heat shock response of murine Chlamydia trachomatis.";
```

```
RL J. Bacteriol. 172:6959-6972(1990).
CC -|- FUNCTION: ACTS AS A CHAPERONE (BY SIMILARITY).
CC -|- DEVELOPMENTAL STAGE: EXPRESSED EARLY DURING INFECTION.
CC -|- INDUCTION: BY STRESS CONDITIONS E.G. HEAT SHOCK (BY SIMILARITY).
CC -|- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE002336; AAF39496.1; ALT_INIT.
DR EMBL: M62819; AAA23138.1; -.
DR HSSP: P04475; IDGA.
DR TIGR: TC0675; -.
DR InterPro: IPR001023; HSP70.
DR Pfam: PF00012; HSP70; 1.
DR PRINTS: PR00301; HEATSHOCK70.
DR PROSITE: PS00297; HSP70_1; 1.
DR PROSITE: PS00329; HSP70_2; 1.
DR PROSITE: PS1036; HSP70_3; FALSE_NEG.
KW Chaperone; ATP-binding; Heat shock; Membrane; Complete proteome.
FT INIT_MET 0 187 0 BY SIMILARITY.
FT CONFLICT 187 187 G -> A (IN REF. 2).
SQ SEQUENCE 654 AA; 70440 MW; EF20C93FE4D3771 CRC64;

Query Match      8.6%; Score 67; DB 1; Length 654;
Best Local Similarity 24.1%; Pred. No. 9.5;
Matches 19; Conservative 19; Mismatches 41; Indels 0; Gaps 0;

QY 39 ELKQAIEDKEATTATEAASSDALEALADOTDALQSEAAVVKADNAASDALEALADOTD 98
DB 567 EKVRAVKEDASTTAIKASDELSARMQKIGEMQAQASANAAGGPINSEDLKKHSF 626

QY 99 ALQSEAEVQVQSDNAASDA 117
DB 627 STRPPAGDNSSSTDNIEDA 645

Search completed: October 13, 2002, 02:09:05
Job time : 27.9572 secs
```

```

QY 114 ASDAWKAATPIALDVYKTKDKTPVVKKE 142
   :  ||| | : | : | :
Db 449 IRRNAKARTDYELKSLKYOEELAQYKKD 477

RESULT 38
YJFJ_ECOLI
ID YJFJ_ECOLI STANDARD; PRT; 232 AA.
AC P39292;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein yjfj precursor.
GN YJFJ OR B4182.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP
SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE-95334362; PubMed-7610040;
RA Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,
RA Blattner F.R.;
RT "Analysis of the Escherichia coli genome VI: DNA sequence of the
RT region from 92.8 through 100 minutes.";
RL Nucleic Acids Res. 23:2105-2119(1995).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U14003; AAC97078.1; -
DR EMBL; AE000490; AAC77139.1; -
DR EcoGene; EG12485; yjfj.
KW Hypothetical protein; Signal; Complete proteome.
FT SIGNAL 1 18 HYPOHETICAL PROTEIN YJFJ.
FT CHAIN 19 232 HYPOHETICAL PROTEIN YJFJ.
SQ SEQUENCE 232 AA; 25333 MW; 65EB1483D6D29369 CRC64;

Query Match 8.8%; Score 68; DB 1; Length 232;
Best Local Similarity 25.0%; Pred. No. 2.5;
Matches 17; Conservative 14; Mismatches 37; Indels 0; Gaps 0;

QY 39 ELKKQAIEDKEATATAEASSDALEALADQTDALQSEEAAYVVKADNAASDALEALADQTD 98
   | : : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 115 EVSRDGVKRAVTATAQRIAQFEQOMEYVVKATEAMQRAQQAQVTTSTVGCASSSVSTAASLK 174

QY 99 ALQSEAE 106
   | : : | : |
Db 175 RIATQRAE 182

RESULT 39
DP3X_ECOLI
ID DP3X_ECOLI STANDARD; PRT; 643 AA.
AC P06710; Q47721;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA polymerase III subunit tau (EC 2.7.7.7) (Contains: DNA polymerase
DE III subunit gamma).
GN DNAX OR DNAX OR DNAX OR B0470.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP
SEQUENCE FROM N.A.
RC STRAIN-K12 / JM109;
RX MEDLINE-87040775; PubMed-3534795;
RA Flower A.M., McHenry C.S.;
RT "The adjacent dnaX and dnaX genes of Escherichia coli are contained
RT within one continuous open reading frame.";
RL Nucleic Acids Res. 14:8091-8101(1986).
RN [2]
RP
SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE-86312915; PubMed-3018672;
RA Yin K.-C., Blinkowa A.L., Walker J.R.;
RT "Nucleotide sequence of the Escherichia coli replication gene dnaX.";
RL Nucleic Acids Res. 14:6541-6549(1986).
RN [3]
RP
SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE-97426617; PubMed-9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Coliade-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [4]
RP
SEQUENCE FROM N.A.
RC Roberts D., Allen E., Araujo R., Aparicio A., Chung E., Davis K.,
RA Duncan M., Federspiel N., Hyman R., Kalman S., Komp C., Kurdi O.,
RA Lew H., Lin D., Namath A., Oefner P., Schramm S., Davis R.W.;
RN Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
RN [5]
RP
RIBOSOMAL FRAMESHIFTING.
RX MEDLINE-90245360; PubMed-2186364;
RA Blinkowa A.L., Walker J.R.;
RT "Programmed ribosomal frameshifting generates the Escherichia coli
RT DNA polymerase III gamma subunit from within the tau subunit reading
RT frame.";
RL Nucleic Acids Res. 18:1725-1729(1990).
RN [6]
RP
RIBOSOMAL FRAMESHIFTING.
RX MEDLINE-90251627; PubMed-2187190;
RA Flower A.M., McHenry C.S.;
RT "The gamma subunit of DNA polymerase III holoenzyme of Escherichia
RT coli is produced by ribosomal frameshifting.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:3713-3717(1990).
RN [7]
RP
RIBOSOMAL FRAMESHIFTING.
RX MEDLINE-90207226; PubMed-2181440;
RA Tsuchihashi Z., Kornberg A.;
RT "Translational frameshifting generates the gamma subunit of DNA
RT polymerase III holoenzyme.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:2516-2520(1990).
RN [8]
RP
REVIEW.
RX MEDLINE-92246902; PubMed-1575709;
RA O'Donnell M.;
RT "Accessory protein function in the DNA polymerase III holoenzyme from
RT E. coli.";
RL Bioessays 14:105-111(1992).
CC -!- FUNCTION: DNA POLYMERASE III IS A COMPLEX, MULTICHAIN ENZYME
CC RESPONSIBLE FOR MOST OF THE REPLICATIVE SYNTHESIS IN BACTERIA.
CC THIS DNA POLYMERASE ALSO EXHIBITS 3' TO 5' EXONUCLEASE ACTIVITY.
CC -!- FUNCTION: THE TAU CHAIN SERVES AS A SCAFFOLD TO HELP IN THE
CC DIMERIZATION OF THE CORE COMPLEX.
CC -!- FUNCTION: THE GAMMA CHAIN SEEMS TO INTERACT WITH THE DELTA
CC SUBUNIT TO TRANSFER THE BETA SUBUNIT ON THE DNA.
CC -!- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate - N diphosphate
CC + [DNA](N).
CC -!- SUBUNIT: CONTAINS A CORE (COMPOSED OF ALPHA, EPSILON, AND THETA
CC CHAINS) THAT ASSOCIATES WITH A TAU SUBUNIT WHICH ALLOW THE CORE
CC DIMERIZATION TO FORM THE POLIIV' COMPLEX. POLIIV' ASSOCIATES WITH
CC THE GAMMA COMPLEX (COMPOSED OF CHAINS GAMMA, DELTA, DELTA', PSI,
CC AND CHI) AND WITH THE BETA CHAIN. THE FINAL COMPOSITION OF THE
CC COMPLEX IS: (ALPHA,EPSILON,THETA)[2]-(TAU[2]-(GAMMA,DELTA,DELTA',
CC PSI,CHI)[2]-BETA[4]).

```



RESULT 35

